INVESTIGATION OF THE STATISTICS OF OCEAN CURRENT SPEEDS

Lorrence Alger Mahaffy

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

INVESTIGATION OF THE STATISTICS
OF
OCEAN CURRENT SPEEDS

by

Lorrence Alger Mahaffy, Jr.

September 1974

Thesis Advisor: R. G. Paquette

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Investigation of the Statistics of Ocean Current Speeds		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis; September 1974 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Lorrence Alger Mahaffy, Jr.		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE September 1974 13. NUMBER OF PAGES 82
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ocean Current Speeds	Current Distributions
Current Velocities	Ocean Current Statistics
Currents	Drift-of-Ship Statistics
Ocean Currents	Drift-of-Ship Speeds
Current Meter Data	Drift-of-Ship Data

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

An in-depth study of 29 time-series current-meter records shows that the logarithmic-speed distributions as a group can be considered to plot symmetrically about their mean, and that the distribution of the mean appears log-normal. The mean distribution does, however, exhibit a slight systematic deviation possibly due to transient phenomena in the data. Fifty drift-of-ship records from the National

Block #20 Continued

Oceanographic Data Center were examined and found (after a necessary alteration) to show the same distributional characteristics as the current-meter data. Indications from drift-of-ship data were that area and seasonal influences affect the speed variability but not the distributional characteristics of the logarithmic-speed transformation. The log-normal distribution for both current-meter and corrected drift-of-ship data appears to be useful out to a deviation from the mean of between two and three sigma units.

Investigation of the Statistics of Ocean Current Speeds

by

Lorrence Alger Mahaffy, Jr.
Lieutenant Commander, United States Navy
B.S., University of Kansas, 1963

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OCEANOGRAPHY

from the

NAVAL POSTGRADUATE SCHOOL
September 1974

ABSTRACT

An in-depth study of 29 time-series current-meter records shows that the logarithmic-speed distributions as a group can be considered to plot symmetrically about their mean, and that the distribution of the mean appears log-normal. The mean distribution does, however, exhibit a slight systematic deviation possibly due to transient phenomena in the data. Fifty drift-of-ship records from the National Oceanographic Data Center were examined and found (after a necessary data alteration) to show the same distributional characteristics as the current-meter data. Indications from drift-of-ship data were that area and seasonal influences affect the speed variability but not the distributional characteristics of the logarithmic-speed transformation. The log-normal distribution for both current-meter and corrected drift-of-ship data appears to be useful out to a deviation from the mean of between two and three sigma units.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDF	Cumulative Distribution Function
CUE	Coastal Upwelling Experiment
DOS	Drift of Ship
e.c.d.f.	empirical cumulative distribution function
K-S	Kolmogorov-Smirnov
MS	Marsden Square
NDM	Normalized Deviation from the Mean
NODC	National Oceanographic Data Center

ACKNOWLEDGEMENTS

The author wishes to express a warm and sincere appreciation to the many people whose unselfish giving of their time, knowledge and experience were most gratefully received and invaluable during the 603 hours of work expended on this thesis from inception to completion.

The following people in particular made significant contributions towards making this project a success:

1. My wife, Judy, and two sons, Scott and Jon Paul,
2. Associate Professor Robert G. Paquette,
3. Professor Donald P. Gaver, Jr.,
4. Second-shift supervisor, Edwin V. Donnellan and his assistants at the W. R. Church Computer Center.

I. INTRODUCTION

A. GENERAL

The broad general subject area of ocean currents has been the recipient of increasing investigational efforts and monetary expenditures. Research has encompassed the spectrum of ocean-current related subjects from the enormous task of determining the effects should a whole current system (Gulf Stream, for instance) be diverted, to studying the effects currents in the ocean have on the growth and decay of density/salinity microstructure in specific localities. A sound understanding of all facets of ocean currents will no doubt prove to be a large step forward in completing man's knowledge of the oceans. One facet of ocean currents which has received limited attention in research studies is the statistical properties of measured ocean-current speeds. This is the specific subject area covered in this report.

B. MEASUREMENT OF CURRENT VELOCITIES AND SUBSEQUENT STATISTICAL STUDIES

1. Ways of Measuring Ocean-Current Velocities

Two means have existed for directly determining current velocities. One has been to place a stationary or semi-stationary device in the water which recorded the flow speed of water around the device. The second way has been to record the set and drift of an object placed in the current. Nearly all reported investigations in which statistical

procedures were used appeared to have based their analyses on time-series current-meter data.

2. Categories of Statistical Studies

A statistical approach to measured current velocities has been the subject or subsection of reported investigational efforts in Russia, Canada, France, Norway and the United States. It appears that statistical procedures, as applied to ocean currents, can be divided into two categories; those dealing in the study of the spectra of ocean currents, and those dealing with the distributional aspect of current velocities. Many of the studies have been concerned with relating spectral properties of ocean currents to internal waves, planetary waves, and theories of turbulence. Others, to a great extent, ignore the spectral properties of a set of data and are concerned with the distributional properties of velocity components and speed values.

3. Current-Speed Statistical Studies

Russia and the United States have published the majority of the reports concerned with the statistical distribution and analysis of ocean-current velocities. Webster [Ref. 1] described and discussed some elementary operations and data presentation techniques for the analysis of a long time-series of current-meter observations. Belyayev and Ozmidov [Ref. 2], using data measured at a semipermanent buoy station, derived empirical distributions of the current-velocity components at ten depths, from 25 to 1200m. It was shown that these distributions differed substantially from

normal below the pycnocline and that the third and fourth moments of the distributions changed abruptly in the pycnocline.

Paquette [Ref. 3] concentrated his efforts on the speeds of current-meter records. He showed that in nearly 80% of the time-series current-meter records checked, when the number of occurrences is plotted against the logarithm of the speed to base ten, the typically skewed distribution of speed becomes Gaussian at the 0.05 level of significance or greater. On long time-series records, the logarithmic standard deviation appeared to range between 0.15 and 0.32. He also concluded that part of the distortion often observed in the tails of the probability distribution of the data was presumably due to inherent current-meter errors. Paquette's results concerning the distribution of the data were presented on cumulative probability plots on which the empirical distribution of the data was plotted along with a normal distribution. In general the appearance of the empirical distribution was quite close to normal, and when subjected to a Kolmogorov-Smirnov (K-S) test for normality, the results suggested this to be true. However, Paquette did not analyze the results further to show whether, on the average, the logarithmic speed distribution produced a nearly normal curve or some distribution close to normal but with systematic deviations from normality.

Paquette briefly introduced and analyzed a limited amount of drift-of-ship (DOS) data. The results indicated

that this type data compared favorably with the majority of the current-meter data. However, since DOS data was not extensively analyzed, the results were not firm and no comparison was made between moored current-meter data and DOS data.

C. PURPOSE

The purpose of this paper is to investigate more closely the normality of the logarithmic-speed distribution as it is applied to ocean-current records and to analyze more extensively DOS data. It will be shown that DOS data (after a necessary alteration) and current-meter data compare quite favorably and that the logarithmic-speed distributions of both types of data can be considered to be symmetrically dispersed about their means with a high level of confidence. The mean value of a group of logarithmic-speed distributions is shown to have a slight systematic deviation probably due to transient phenomena.

In order to extend the studies of current-meter time-series data to a new area of the ocean and to add more samples to the data base, eleven current-meter records from the Coastal Upwelling Experiment (CUE) off the coast of Oregon were also analyzed.

The basic approach in this presentation and analysis of results has been two-fold. One was to record and plot, at normalized deviations from the mean (NDM) of each cumulative probability distribution, the difference value between the logarithmic distribution and a log-normal distribution. If

an empirical logarithmic distribution were truly normal, the differences would be zero and a straight line through the zero values of the plot would occur. The second procedure was to apply known statistical measures and interrelationships to parameters derived from the first four moments of a distribution. Specifically these parameters are the mean, standard deviation, coefficient of skewness, and coefficient of kurtosis.

II. THE DATA

Data used in the statistical analysis and relationship investigations reported in this paper came from four sources.

A. TIME SERIES DATA

1. Sources of the Data

One source was moored current-meter data recorded by Woods Hole Oceanographic Institution [Refs. 4, 5, 6]. The second source was moored current-meter data recorded by Paquette and designated SCARF 1 through SCARF 7. These first two sources include 29 of the 43 time-series records used by Paquette [Ref. 3].

A third source was eleven sets of moored current-meter records furnished by Donald Bishop in the office of the Coastal Upwelling Experiment (CUE) at the University of Washington. This data was recorded by Oregon State University at a rate of one speed record every 5 or 10 minutes. TABLE I gives the basic statistical summary of the CUE data. In reference to TABLE I, the sample identification number provides an indication of the meter's location (these locations are shown in Fig. 1). $\overline{V}$ is the arithmetic mean of the speed. $\overline{\text{Log V}}$ gives the mean of the logarithmic-speed distribution. σ is the arithmetic standard deviation while σ_L stands for the standard deviation of the logarithmic-speed distribution. ΔPm is the maximum difference observed between the cumulative probability of the logarithmic-speed distribution and the cumulative probability of a log-normal

distribution over the same speed range. P gives the cumulative probability of the empirical distribution at the point where ΔPm occurred.

2. Independence of Observations

Time-series data produces questions as to the independence between consecutive data points since most statistical procedures are based on the independency of individual data samples. Time-series data recorded at short intervals are usually autocorrelated which lowers the degree of independence between data observations. The CUE data apparently are highly autocorrelated. The autocorrelation coefficient drops to 0.3 when using one observation every four hours. However, the effects of decimation of the data were not investigated. Paquette [Ref. 3] assumed that the number of effective individual data points (needed for goodness-of-fit tests) in the distributions he used could be obtained by dividing the total number of observations by the number of lags to get to an autocorrelation coefficient of 0.3. He showed that decimination to this degree had negligible effect on the mean and standard deviation of the distribution. It will be assumed that the same procedure can be followed with the CUE data.

B. DRIFT-OF-SHIP DATA

1. Source of the Data

The fourth source of data, and one not extensively utilized by Paquette, came from the files of the National Oceanographic Data Center (NODC) from their File H1-9. This

file is an extensive set of comparisons of dead-reckoning positions and corresponding fixes covering the period 1904-1945. The difference between the dead-reckoning and celestial or electronic fix is ascribed to a current which is presumed constant over the hours and the many tens of miles between fixes. NODC furnished computer-generated printouts which included all information for Marsden Squares (MS) 114, 115, 116, 149, 150 and 151. Their locations are shown in Fig. 2. Selected data from these printouts were used in this analysis. Also shown in Fig. 2 were the basic locations of the Woods Hole current meters whose data were used both by Paquette and in this thesis. The DOS data was reported by five-degree quadrants within each ten-degree Marsden Square, Fig. 3, and then by month, general current direction, and speed interval within each quadrant, Fig. 4.

2. Independence of Observations

DOS data has been computed and reported by an uncountable number of people. The time of day the reports were made, the types of ships involved, the location of each ship, the wind and weather conditions were all unknown factors which were assumed to have encompassed all possibilities over the 41 year reporting period. It is known that DOS data was not recorded when the reported wind speed exceeded Beaufort 7 or seas exceeded 3.3m. With all these factors in mind, it was assumed the DOS data represented basically random independent samples in the areas from which information was reported. This did not exclude the possibility that

peculiarities in the speed classes may exist for various reasons such as errors in grouping the data, bias factors on the part of the reporting navigators, or the kind of space and time averaging involved.

III. COMPUTER PROGRAMS

All the statistical parameters generated from the data used in this paper were obtained using computer programs on the Naval Postgraduate School IBM 360/67 digital computer. Table II provides a summary of the major programs utilized. Minor programs were written by the author to perform specific tasks throughout the course of the investigation but these did not compute statistical parameters.

Program HISTG classifies current-speed data into class intervals and plots the resulting histogram on the line printer. This program was used to generate statistics on the OSU current-meter data, which was received on tape as individual records, and on data sets keypunched on to computer cards. CUDIS MOD3 and CURST2 accept data in histogram form grouped both in even and uneven intervals. CUDIS MOD3 computes statistical information based on the assumption that the number of counts in each speed-class interval are concentrated at the center of the interval, and produces a cumulative log-normal distribution and plots it on a probability-paper scale. CURST2 computes statistical information based on the assumption that the number of counts in each speed-class interval are evenly distributed across the width of the interval. It does not produce a plot. Besides the information provided in Table II, CURST2 computes the third and fourth moments of a distribution and the coefficients of skewness and kurtosis. These are not generated by CUDIS MOD3.

IV. MOORED CURRENT-METER DATA

A. ANALYSIS APPROACH USED

Paquette [Ref. 3] concluded that the current-speed distributions were log-normal at a level-of-significance of 0.05 or greater by testing each of the 43 series studied with the K-S statistic. He used the mean and standard deviation obtained from the data as estimates of these parameters for the parent population. However the K-S statistic Paquette used assumes the parameters of the parent population are not estimated from the data. According to Lilliefors [Ref. 7], when the parameters of the parent distribution are estimated from the data, the probability of a type I error will be significantly smaller than as given by tables of the K-S statistic. Lilliefors provides a new table for the critical values of the deviation for several useful α values. The values used to construct Fig. 5 were obtained from this table. The effective number of observations is along the abscissa with the maximum permissible deviation values plotted on the ordinate. Thus the results obtained by Paquette are conservative in that his results are at a higher level-of-significance than they should be.

All current-speed data used by Paquette [Ref. 3] and in this thesis were generally is histogram form. It is realized the K-S test was derived for ungrouped data and that its behavior is less well understood when using grouped data. Current-meter data were grouped in only one cm/sec intervals

and appeared more or less as continuous data. However, DOS data was highly grouped and less acceptable for application of the K-S statistic. Therefore, it was assumed that the DOS data would give a larger maximum difference between cumulative distributions than ungrouped data (an assumption which seemed reasonable), and that the K-S test would give a reasonable result that was somewhat liberal (reject more than it would if the data were not grouped). More work on this subject is needed but is left for future studies.

However, if the current-speed distributions are in general log-normal, one might assume that the normalized-logarithmic distributions derived from the many time series ought to be comparable and members of an ensemble of distributions. Then one may test the fit by examining the deviations of the cumulative distribution function (C.D.F.) of the data from the cumulative log-normal distribution at a number of values of the normalized deviation of the logarithmic speed, $\frac{(\mathrm{Log}\ V - \overline{\mathrm{Log} V})}{\sigma_L}$, where Log V is the logarithm to the base ten of any speed value V, $\overline{\mathrm{Log}\ V}$ is the mean of the logarithmic-speed distribution, and σ_L is the standard deviation of this distribution. This technique has the advantage of examining all of the series together, looking for an overall systematic difference from the ideal and looking at the distribution of the difference values at each normalized deviation point selected.

Besides the goodness-of-fit to the normal cumulative distribution curve, the coefficients of skewness and kurtosis

were examined as they relate to each other on a Pearson diagram. These coefficients also might be expected to be comparable if the curves are similar. However, as pointed out by Pearson [Ref. 8], different distributions can have the same first four moments. These coefficients apparently have not been used previously in studies of currents.

B. DEGREE OF NORMALITY OF TIME-SERIES LOGARITHMIC-SPEED DATA

1. Data Used and Presentation Methods

The data used in this approach was part of the same time-series data used by Paquette and included all the SCARF data and all the Webster and Fofonoff data, 29 time-series data sets in all. Figure 6 is a plot of the difference values (observed minus predicted logarithmic cumulative probabilities) for the 29 time-series data sets at nine normalized deviations from the mean (NDM). Difference values are noted along the absicssa while the nine NDM values selected are indicated along the ordinate. Plus and minus three sigma units were used as the limits of the NDM values because the time-series records did not provide sufficient values for analysis beyond these points. Bar plots of the difference values at each NDM are given showing the range of values observed. A smooth curve was faired through the mean value at each NDM considered.

Table III provides summary statistics of the data used to construct Fig. 6. Not all of the 29 time series data sets extended out to the two and three sigma location. The last column of this table provides the results of a K-S

goodness-of-fit test for normality of the difference values at each NDM assuming a mean value of zero. Under the hypotheses that the log-speed transformation produces a normal distribution from current-speed records, it is assumed that difference values at each NDM are random and come from a nearly normal population whose mean is zero.

It is readily apparent in Fig. 6 that the range of difference values includes the zero value in all instances. However, the distributions of the difference values are not in general symmetric about zero. This is not too surprising since any subsample drawn from a parent population will most likely not possess the same mean as the parent population. A smooth curve through the mean values at each NDM shows a systematic "S" shape variation from the log-normal curve.

2. Significance of Observed Results

In order to determine the significance of the "S" shape variation in Fig. 6, one must examine some of the statistical values provided in Table III. To aid in this examination, Table IV is given which shows some of the computations and values required in the following analysis. Columns 1, 2, 3, 6 and 7 of Table IV are repeated from Table III. Brooks and Carruthers [Ref. 9] provide computations for the standard error of the coefficient of skewness of any series of N random numbers (p. 55), and the standard error of a single observation from a sample of N observations (p. 40). "t" in Table IV is the value for a "Student" t-distribution, ν is the degrees-of-freedom for that distribution, and α is the

significance levels obtained when entering "Student" t-tables for a two-tailed level-of-significance test. The usage of these values will be explained later.

a. Symmetry of the Data at Each NDM

It is noted in Table IV that at all but one of the nine NDM's, the coefficients of skewness of the individual sets of difference values is less than one. Brooks and Carruthers [Ref. 9] point out that any set of N random numbers will show a certain amount of skewness (p.55), however, the absolute value of the coefficient of skewness less than one indicates data only moderately skewed (p. 56). They also specify that the skewness can be considered real only when the coefficient of skewness exceeds twice the value of the standard error (p. 55). It seems only logical that the larger the value of N, the better the confidence in these statements. By comparison of columns three and five in Table IV, it is seen that except for the NDM value of three sigma, the majority of the coefficients of skewness fall significantly short of being equal to twice the value of the standard error. Since skewness is an indication of the symmetry of a distribution about its mean, the indication from Table IV is that at each NDM except three sigma, the difference values are basically symmetrical about their means.

Since the data at the NDM values are basically symmetrical about their means and since a review of the histograms of the data show in general normal type distributions, except at three sigma where the data exhibits a

definite "J" shaped distribution to the left, a test was made to determine to what degree the data were normal about the hypothesized mean of zero. A K-S test was conducted with the standard deviation estimated from the data. The results are given in the last column of Table III. The figures given are the level of significance or α values of the test obtained from Fig. 5. The amount of reduction in the α value due to estimating only the standard deviation from the data was not known, but it was assumed to be significant and therefore Lilliefor's results were used. It appears the normal hypothesis could be rejected on the basis of the evidence from these data, at a significance level of .008 or below.

b. Significance of the Deviations of the Means

As stated previously, it is a known fact that any subsample from a large population will most likely not have the same mean as the parent population. Brooks and Carruthers [Ref. 9, p. 65] demonstrate a method of testing whether a mean M from a subsample differs significantly from a postulated population mean M'. The test can be made using the well-known "Student" t-distribution where the t-value is computed by:

$$t = \frac{(M - M')}{\sigma/\sqrt{N}} .$$

σ is the estimate of the population standard deviation derived from the sample, and the distribution of "t" is associated with N-1 degrees-of-freedom. This fact can be used

to test the significance of the deviation of the mean from zero at the individual NDM's. At each of the NDM's the value of M' is zero, and the values for the other computations to derive "t" are given in Table IV. The test hypothesis is that the subsample mean does not differ significantly from zero.

Figure 7 is derived from the t-tables and can be used for the t-tests in this thesis. The "Student" t-value is given along the abscissa with level of significance on the ordinate. The curves are for different values of degrees-of-freedom. Enter with the t-value and degrees-of-freedom and read off α on the ordinate.

We can infer therefore from the results in Table IV that the departure from the mean of zero at each NDM value is probably significant except possibly at NDM values of -2 and 0.5. These latter two means are near zero anyway and are near crossing points in the curve. Therefore the "S" shaped curve in Fig. 6 is indeed most probably real, and not a sampling artifact.

c. An Engineering Viewpoint of the Significance of Results

Perhaps more important than significance in terms of probability is the utility of this information from an engineering viewpoint. If one is concerned with the maximum current to be expected on an object being placed in the ocean, he is concerned with the high speed tails of the distribution being correct. At a NDM of two sigma the normal probability ought to be 0.9772. The maximum difference value

observed from the data was 0.023. This gives an error, $\left(\frac{1 - .9772}{.023}\right)$, that is not quite as great as one times the residual probability remaining. At the NDM of three sigma this error increases to over sixteen times the residual probability remaining. Therefore, the data at the three sigma value is unreliable for use, as will be shown below.

d. General Summary and Possible Errors

It can be said that the "S" shape curve in Fig. 6 is most probably real as shown and not due to chance. In general the type of curve represented by the smooth curve in Fig. 6 is one which contains slightly fewer data points below the mean and slightly more data points above the mean than would be expected of normally distributed data. To put Fig. 6 into a better perspective to indicate just how much deviation is being shown, a more familiar representation of the CDF's of the curves in question is shown in Fig. 8. As can be seen, the maximum deviations are small and the CDF's of the two distributions are almost identical.

Perhaps one could argue that the systematic deviation in Fig. 6 could be caused by measurement errors or errors in treating the data. This could possibly be true if it were not for the fact that the data used for Fig. 6 came from two separate sources and that a similar plot using only the eleven sets of CUE data (which is yet a third source and one which used a different type of current meter, the Aanderaa, than the SCARF and Webster and Fofonoff data) showed the same general variation. Therefore the reason for

this systematic variation is not clear. Some possible causes could be the occurrence of events such as storms which may produce anomalous water velocities for a substantial fraction of the recording period, mooring transits, and excessive oscillation of the buoy during the recording period. All of these could conceivably produce the type of effect noticed.

e. Limit of Usefulness

It was shown that at all NDM's, the difference values obtained were basically symmetrical about their mean and the histograms indicated possibilities of normality except at the three sigma location. The distribution here was "J" shaped trailing off to the left or towards higher negative difference values. This says that at a NDM of three there is in general fewer observations than observed in a normal curve of the data. This is to be expected since the current meters have a tendency to record fewer than observed higher speeds due to a coalescing of speed dots on the recording film. Therefore it appears the NDM value of three is beyond the usefulness of the current meter to provide satisfactory data for analysis. Although current meters do have problems with stalling of the rotor at the low-speed end of the scale, the data at a NDM value of minus three does not indicate any problems, so it is assumed this value is within the useful range of the current meters used. Since data obtained for very high and very low speeds is suspect, no explicit attention has been paid to this in the process of statistical estimation. New "robust" procedures that account for such data difficulties are described by Andrews [Ref. 10].

C. PEARSON DIAGRAM

1. Presentation Method

Another means of determining the type of distribution data may represent is by use of a Pearson diagram. This is a diagram on which is plotted the square of the coefficient of skewness, β_1, versus the coefficient of kurtosis, β_2. Pearson [Ref. 11] showed that different regions of the β_1, β_2 space correspond to several different theoretical distribution curves. Table V provides the β_1 and β_2 values for the logarithmic time-series data previously considered plus these values for the CUE data which will now be included for analysis. Figure 9 is a Pearson diagram on which the β_1, β_2 values from Table V are plotted.

2. Indication of Data Errors

The plotted points in Fig. 9 appear to show an excessive spread. However, further investigation into comments concerning the recording of the SCARF and Webster and Fofonoff data showed that about 63% of the data sets having a β_2 value of 4.5 or below experienced marked quantization in speeds, higher than normal speeds due to mooring transits, or excessive buoy oscillations while in place. About 67% of the distributions with values of β_1 equal to 1.0 or greater showed these same characteristics. Only one of the CUE data records plotted in the region just discussed but no detailed information on those records was readily available.

One of the errors mentioned above, high speeds due to mooring transits, does add a quantity of high speed values

to a speed record. Transient phenomena such as storms and influences from high speed current regimes can also cause an excess of high speed current values. These high values cause the distribution to be more positively skewed than otherwise would be expected. These factors could distort a speed record significantly if the total recording time is small. Many of the records with large β_1, β_2 values were also relatively short time duration (less than a day). Pearson [Ref. 8, p. 285] discusses this problem of long tails on a distribution and shows that the contribution to moments from the tails significantly increases as the moment increases. For instance, the contribution to the fourth moment from areas in the outer .001 part of the tail, of a distribution with β_1, β_2 values of 2.79 and 9.01 respectively, is about 41%. This contribution increases to 74.2% if the outer .01 part of the tail is considered. Since the β_1 and β_2 values depend on the second, third and fourth moments, erroneous speed values which extend the tails of a distribution will have significant effect on where a distribution plots on a Pearson diagram.

It would be impossible, without a highly detailed study, to ascertain to what extent the three errors mentioned influenced the data, but is fairly obvious that some had significant influences on the high values of β_1 and β_2 observed in Fig. 9. None of these problems were noted in data which exhibited β_1, β_2 values smaller than the values given above.

3. Summary of Results

A normal curve will generate β_1 and β_2 values of 0.0 and 3.0 respectively. A grouping of points about the (0,3) value would be an indication the log speed transformation was a good fit. Except for the points previously mentioned, the majority of the logarithmic time-series data plots closely grouped about the (0,3) point. Again the indication is that the log-normal approach to current-meter time-series data produces a near-normal distribution. This diagram will be used later to compare with the DOS data.

D'Agostino and Pearson [Ref. 12] and Bowman [Ref. 13] have published recent articles on the use of the β_1, β_2 statistics in testing normality of a data set. Their procedures were not used in this thesis, but are referenced for future use.

V. DRIFT-OF-SHIP DATA

The attention of the analysis effort then shifted to the DOS data. The number of locations where current meters have recorded measurements is small in comparison to the total area of the ocean. However, DOS information is available over a large percentage of both the Atlantic and Pacific Ocean. If a suitable distribution for these speeds could be found, a method would be available for estimating the speeds probabilistically. This requires also some way of estimating the second moment, a quantity which is not charted on the current charts. Since DOS data are somewhat different and probably more distorted than current-meter data, it is desirable to use current-meter data to help correct the distortions.

It is recognized that ocean currents usualy decrease with depth. This is an important part of the current prediction problem to engineers. The present study does not enter into this problem.

A. IRREGULARITIES IN DOS DATA

DOS data used in this study appear to suffer some irregularities at both ends of the speed spectrum. This is discussed to some degree by Paquette. The four knot speed class (see Fig. 3) includes all accepted speeds four knots and greater. This has the effect of requiring one to place a limit on the upper class interval in order to proceed

with distributional investigations. Herein enters one possibility for error. Although the total number of occurrences in this speed class is small in comparison to the total count, the generated errors could be significant. A speed of 4.5 knots was chosen as the top limit for this analysis.

The lowest class interval also presents a problem. Although described as "calm," its upper boundary is slightly less than 0.1 knot. It is assumed that true zeros do not exist and the lower boundary is placed at 0.01 knot. Small changes in this arbitrary choice have considerable effect when the logarithmic transformation is made. Furthermore, after transformation the class interval is too large to properly represent the tail of the curve. A pictorially nicer technique would be to distribute the counts in this interval into several intervals according to a rule consistent with the log-normal curve. This seemed like too much tampering with the data and the above simple course was followed.

It is apparent that wind effect included in the recorded speeds is impossible to ascertain. It could add to or reduce from the true current speed. This would vary with wind speed, direction of ship travel relative to the wind, and from ship to ship. No wind correction factors were entered. As was mentioned, data taken when the winds were above Force 7 are excluded from the data. While this reduces the effect of excessive wind-drift of the ship, it also eliminates the higher speeds of wind-driven current.

It is also to be noted that the DOS data are averages in time and space. This averaging will smooth sharp high and low peaks and will reduce the apparent numbers, especially of the high speeds.

Human error certainly enters into the results. In most cases one expects this to be Gaussian error and to have little effect except to increase the standard deviation slightly. However, there appears to be a significant bias at the low-speed end of the curve which will be discussed in the next section.

B. A NECESSARY DATA ALTERATION

There is an apparent anomaly in the "Calms" and 0.1 knot speed classes. It is believed that this is an artifact arising from a natural but unjustified pride in precision of celestial and electronic fixes. There is nearly always some scatter among the navigational lines of position. It would be natural for the navigator to be biased toward those which agreed with the dead-reckoning position. So it would not be surprising to find more recorded "calms" than actually occurred.

1. Alteration Indicated by e.c.d.f.

This appeared to be an explanation for the results observed when the empirical cumulative distribution function (e.c.d.f.) of the DOS data is plotted. The e.c.d.f. is a plot of the i-th ordered value as ordinate against $(i-\frac{1}{2})/N$ as abscissa. N is the total data count. In one-dimensional samples, it provides an exhaustive representation of the data

under the following broad assumptions: (i) that the order of the observations is immaterial, (ii) that there is no classification of the observations, based on extraneous considerations, which one wishes to employ; and (iii) if the sample is non-random, then appropriate weights are specified. Wilk and Gnanadesikan [Ref. 14] discuss the significant advantages of using the e.c.d.f. in a descriptive test of data. It is pointed out by them that the e.c.d.f. "is a robust carrier of information on location, spread and shape, and an effective indicator of peculiarities" (p. 2). Figure 10, included as an example of the type of plot one might expect to see from a log-normal data series exhibiting no readily apparent data irregularities, is the e.c.d.f. for Webster and Fofonoff measurement No. 1012 (WF 1012). One sees basically a smooth flow of the data from one end to the other. Plotted in Figure 11 is the e.c.d.f. of MS 115, quadrant 1, month 10. The data flow appears smooth in the upper 60% of the observations, but some peculiarity is evident in the lower end of the data. It was felt two basic reasons caused this to occur. One is the lack of resolution of speeds in the region near zero. However, despite this factor, it appears likely the main reason is that too many observations occur in the "calm" class. If some were transferred to the 0.1 knot speed class, the e.c.d.f. plot would appear smoother.

Plots of the e.c.d.f. of other sets of DOS data showed similar traits to varying degrees. It was not feasible to investigate this characteristic of the data more thoroughly

at this time. Therefore, a partial correction was made by arbitrarily shifting nine-tenths of the counts in the "calm" interval into the 0.1 knot interval. Figure 12 is the e.c.d.f. for the data of Fig. 11 altered in this way. It shows a much smoother data fit and one which generally resembles the current-meter data of Fig. 10, except for the inflection at the lower end which may be obscured by the coarseness of the subdivision into intervals.

2. Alteration Indicated by Probability Density Plot

Visual examination of the log-normal probability-paper plots for the cases studied showed this arbitrary change to be at least approximately correct. As an example, Fig. 13 shows two separate logarithmic probability density plots for MS 116-4-6 and MS 116-3-9. Each include the results of one unchanged data base and one corrected as discussed above. The effect of the nine-tenths shift is very dramatic and produces a more normal appearing probability density plot. No further investigation was done to determine whether the nine-tenths shift was an optimum alteration.

C. DOS STATISTICAL PARAMETERS

1. Data Used and Presentation Methods

Fifty different months of DOS data were selected for analysis from areas generally off the northeast coast of the United States. Each set of data was distributed over a five-degree square. The squares and months were selected to provide data from within and outside areas of expected high current velocities at different times of the year due to major current

systems. These fifty data sets were then altered by the nine-tenths data shift and then analyzed with the aid of the computer programs CUDIS MOD3 and CURST2.

Table VI provides the statistical summary of the data generated by these programs. Columns one and two identify the data sets and indicate the number of speed class intervals into which the data is divided as well as the total speed observations per data set. The arithmetic mean ($\overline{V}$) and standard deviation (σ) are given in columns three and four respectively. Columns five through eight provide the logarithmic statistics for each data set and include in the order given, mean ($\overline{\text{Log V}}$), standard deviation (σ_L), coefficient of skewness and coefficient of kurtosis. As defined before, ΔPm is the maximum difference between the logarithmic-speed cumulative probability and a log-normal cumulative probability, while P is the value of the logarithmic-speed cumulative probability where ΔPm occurs. For comparison, these values are given for the curve that existed prior to the nine-tenths alteration.

2. <u>Comparison With Current-Meter Data</u>

Several results become readily apparent from Table VI when compared with Paquette's work [Ref. 3] and Table I of this report. The logarithmic standard deviation appears to be grouped into narrow limits between 0.24 and 0.36. This measurement for the current-meter data ranged between 0.10 and 0.538. This is attributed to the grouping of the DOS data and the limit placed on the high-speed end of the

DOS data. All the DOS data sets except one show a small to moderate negative skewness. As can be seen from Table V this was generally true for the current-meter data, however some exhibited skewness coefficients that were positive and some that were negative but greater than minus one. The coefficients of kurtosis for the DOS data ranged between 2.49 and 5.358 while for logarithmic current-meter time-series data they ranged in value from 2.46 to 12.91. The two sets of data look generally alike except the current-meter data is considerably more variable. Some deviations in the current-meter results are so extreme that peculiarities in the data are suggested.

3. K-S Test of Normality of Each Data Set

In order to produce a numerical measure of closeness of fit of the logarithmic current-meter distributions to the log-normal, Paquette [Ref. 3] applied the K-S statistic as previously mentioned. The K-S statistic uses the maximum deviation in absolute value between the empirical and theoretical cumulative distribution (ΔPm in Table VI) and the effective number of observations (number of independent observations) to derive a level-of-significance for the fit.

The K-S test was applied to the DOS data in Table VI using Lilliefors' results. The total number of observations in each data set was used to enter Fig. 5. At an α level of 0.05, the maximum permissible deviation was obtained from the ordinate. If this value was greater than ΔPm in Table VI, the normal hypothesis could not be rejected on the basis

of the evidence from this data at the significance level of 0.05. Ninety percent of the DOS data sets passed the K-S test with a confidence level of 0.05 or greater. The same procedure was used to test the data sets prior to the nine-tenths alteration. Nearly 88% of the unaltered DOS data sets failed the K-S test at the 0.05 level of significance. It therefore appears that the nine-tenths data alteration in the first speed class produces a much more normal logarithmic-speed distribution.

D. DEGREE OF NORMALITY OF DOS LOGARITHMIC-SPEED DATA

1. Data Presentation

The same procedures as used with the current-meter data were applied to the DOS data. A bar plot of the difference values between the observed and predicted cumulative distributions at designated NDM values is presented in Fig. 14. The striking resemblance in shape to Fig. 6 is readily apparent. Table VII provides a summary of the statistics from the data used in constructing Fig. 14. This table corresponds to Table III. The distribution of the difference values at the individual NDM's is not entirely symmetric about zero, and a curve smoothed through the mean value at each NDM shows a slight "S" shaped systematic variation from the normal curve.

The distribution represented by the smooth curve through the mean values of the DOS difference values in Fig. 14 is nearly the same as the current meter data except it is more symmetrical in shape than the curve in Fig. 6.

2. Symmetry of the Data at Each NDM and Overall Limits of Data Usefulness

Table VIII is like Table IV except that the figures come from the DOS data under consideration. A review of the coefficients of skewness in Table VIII show that except at the three sigma location, all values are significantly less than unity, indicating only moderate skewness. All but two of the coefficients of skewness are significantly less than twice the standard error, indicating that their skewness is probably not real and the data are nearly symmetric about their means. NDM values of minus one and three showed signs of real skewness in the distribution of difference values.

A visual survey of the histograms of the difference values at each NDM point revealed basically normal looking distributions except at the three-sigma location, where the distribution resembled a "J" shaped curve trailing off to the left toward higher negative values. This indicated fewer observations were observed in this area than expected. This result should be expected since restrictions based on wind force and sea state at the higher-speed end of the data probably eliminated many of these higher-current values. A similar phenomenon was observed with the current-meter data. It is interesting to note that because of the restrictions on the high speed ends of the current values both current-meter and DOS records showed a similar exponential distribution at the three sigma point with nearly identical values for the coefficient of skewness and coefficient of kurtosis. Unless some means is derived to correct for the lost data in

the high speed tail of the DOS distributions such as extending the upper limit of the 4.0 knot speed class, the three sigma point, as with current-meters, appears to limit the useful range of DOS data.

A K-S test was conducted at each NDM to test the hypothesis that the data at these points were normal about the theoretical mean of zero. The results are shown in the right hand column of Table VII. One sees that there is little or no likelihood that the data could be normal about zero. This corresponds to the results obtained from current-meter data as shown in Table III.

3. <u>Significance of Deviations of the Means</u>

A "Student" t-test was made to test the hypothesis that at each NDM, the deviation of the data mean from the theoretical mean of zero is not significant. The computations and results of this test are given in Table VIII. Only two locations passed this test with a level of significance greater than 0.05. These two points, -0.5 sigma and one sigma, are near crossing points of the smooth curve in Fig. 14 and therefore concurrence with the hypothesis would be high at those points. As was found in Fig. 6, the deviations causing the "S" shape curve in Fig. 14 are most likely real.

E. PEARSON DIAGRAM USING DOS DATA

A plot of the β_1, β_2 values for the fifty logarithmic DOS data sets on a Pearson diagram is shown in Fig. 15. No exceedingly large values of β_1 and β_2 were obtained from the DOS distributions and no attempt was made to identify any

irregularities in the two distributions that had a β_2 value greater than 4.5. It is readily apparent the DOS data is closely grouped around the (0,3) point on the diagram and compares most favorably to the majority of the current-meter distributions in Fig. 9. The Pearson diagram has been used as an indication of normality and as a tool for general comparison between the two types of data included in this thesis. The full utilization and subsequent implications one could employ with regard to current-speed distributions through use of the Pearson diagram are left to future work in this area.

F. GENERAL BAR PLOT COMPARISON BETWEEN CURRENT METER AND DOS DATA

Figure 16 is a composite of Fig. 6 and Fig. 14 plotted together for comparison. It is readily apparent that the variability in difference values is more extreme for current-meter data than for DOS data. The most likely reasons for this is that the DOS data are highly grouped, in general contain only a moderate number of observations, and those observations that are available have been averaged over time and space due to the nature of the recording technique. It is possible that the DOS data are a better measure of the statistics of current measurements made over very long periods than are the current-meter data, having gained more from their randomized distribution over years of time than they have lost from their various known distortions. It would seem, therefore, that the difference in variability has little significance.

Another apparent discrepancy in Fig. 16 is that it seems the smooth curve through the means of the current-meter difference values is offset from the DOS curve by about one sigma unit. What probably has happened is that the lower tail of the DOS data has been compressed towards the upper tail by about one sigma unit. This also accounts for the fact that no DOS data sets exhibited low-speed values which extended out to three standard deviations from the mean. The reason is that the grouping into class intervals centers the speed of the lowest class interval higher than the several low-speed class intervals in the current-meter distribution. These low speeds become relatively large deviants from the mean after the logarithmic transformation. It was necessary to make an *ad hoc* readjustment of the DOS data in the low-speed end while at the same time setting a rigid boundary on the high-speed end. No such adjustments were necessary for the current-meter data.

G. OTHER POSSIBLE VARIATIONS WITHIN DOS DATA

Some of the variability noted in the current-speed data used in this report could have come from differences in area influences on the data and differences in seasonal influences on the data. Because the DOS data was available in a large quantity covering both area and time, an effort was made to check for possible indications of these two types of variability using the DOS data only. The procedure was to select the data and plot it in the bar format similar to Fig. 6. The number of distributions that could be used from the

fifty DOS data sets previously studied ranged between eight and ten for each case cited below. Because the data base was small, the plots generated were used to provide possible indications of differences without proceeding further with significance tests or in-depth reasoning for their existence.

1. Area Influence

Figure 17 is a plot using data from two separate areas, MS 149-3 and MS 114-1, to check for possible area influence on current speeds. Individual months in each area were taken as separate data sets. The plot indicates the surface current speeds in MS 149-3 are in general more variable over a year's period than in MS 114-1. This is not too surprising since MS 149-3 is east of Newfoundland and probably more susceptible to storms and current variations; the area is located in the vicinity where the Labrador and Gulf Stream current regimes generally mix. MS 114-1 is in the mid-Atlantic east of Charleston, South Carolina, where active and variable current-speed conditions are not known to exist. However, the general shape of the two curves is about the same. Therefore, the indication is that possibly an area influence on surface current speeds exists affecting the variability of the speeds but not the general shape of the distribution of the logarithmic-speed curve.

2. Seasonal Influence

Figure 18 is a plot using data from two separate seasons of the year, winter and summer. For winter the month of January was selected and the data randomly covered all MS areas except MS 148. For summer, the month of July was selected

and the data covered the same MS areas and quadrants from which the January data was taken. The plot indicates that in the summer the currents are in general more variable in magnitude than the winter currents, however the general shape of the curves is somewhat similar. The implications of the variability noted cannot be readily related to any current regimes. Since the factors which create and maintain ocean currents are numerous and sometimes unpredictable, an in-depth study would be needed to confirm these variability results and then to establish reasons for their existence. However, there are indications that seasons do influence the variability but not the distribution of DOS current-speed data.

VI. CONCLUSIONS

The logarithmic-speed transformation of both current-meter and altered DOS speed records produces distributions that as a group can be considered symmetrically distributed about their means. The distribution of the mean values appear log-normal. However, the mean of both types of data exhibit a slight "S" shape systematic deviation which is probably real. This systematic deviation appears likely to be the result of external influences on the data which are both natural and man-made. Indications are that elimination of these influences would allow the mean of the logarithmic-speed distributions of current data to be log-normal with a high level of confidence.

DOS data compares quite favorably to current-meter data and could be used to derive probability estimates for surface current speeds in areas where no other data is available.

The limits of usefulness of current-meter data appears to extend from NDM values of at least -3 sigma to somewhere between two and three sigma. For DOS data these limits are from at least -2 sigma to somewhere between two and three sigma. Extrapolation beyond these limits is extremely uncertain and, with present knowledge, should only be considered if the consequences of a many-fold error in probability are freely accepted.

Indications are that seasonal and area influences on the current speeds exist but that these influences are limited

Table I. Statistical Summary of Moored Current-Meter Records from Oregon State University's Coastal Upwelling Experiment.

SAMPLE IDENTIFICATION NUMBER		RECORDING DATES		# OF SAMPLES	DEPTH OF METER (m)	$\overline{V}$ cm/sec	$\overline{\text{Log } V}$	σ cm/sec	σ_L	ΔPm	P
		FROM	TO								
NH-6	491/8	8/31/72	10/29/72	8486	25	20.60	1.29	8.13	.145	.0442	.8203
NH-6	490/8	8/31/72	10/29/72	8486	40	17.60	1.22	6.91	.151	.0526	.7761
NH-15	455/5	4/17/72	5/18/72	8858	20	39.46	1.58	9.09	.110	-.0670	.2697
NH-15	456/5	4/17/72	5/18/72	8858	80	20.05	1.28	6.45	.150	-.0742	.5503
NH-15	453/10	5/18/72	5/31/72	3975	20	27.21	1.42	7.37	.124	-.0624	.5648
NH-15	452/7	5/18/72	5/31/72	3974	80	14.23	1.12	6.81	.174	-.0329	.5964
NH-15	455/10	5/31/72	6/20/72	5764	20	20.16	1.29	5.38	.121	-.0410	.4519
NH-15	456/10	5/31/72	6/20/72	5600	80	13.09	1.09	4.96	.163	.0495	.4336
NH-15	D72/7	6/20/72	7/18/72	7700	0 (SURF)	5.26	.66	2.54	.257	-.0865	.5058
NH-15	454/12	6/20/72	7/18/72	7700	20	20.72	1.28	7.90	.177	-.0575	.3670
NH-15	452/10	6/20/72	7/18/72	7700	40	16.20	1.18	5.35	.153	-.0644	.6003

Table II. Summary of Major Computer Programs Utilized

PROGRAM NAME	PROGRAMMER	PURPOSE/COMMENTS
CUDIS MOD 3	R. G. Paquette Assoc. Professor Naval Postgraduate School, Monterey	To compute log-normal statistics on histogram data when the class intervals are variable or constant. For input it requires the center and top speed for each interval (number of intervals restricted by computer stowage) and the counts per interval. It's output includes a printer plot histogram of the input data and tables of arithmetic and log-normal statistics including the difference between the CDF of the data and the CDF of a normal distribution. It also provides for a single CALCOMP plot of the arithmetic, logarithmic and normal cumulative probability curves.
CURST 2	R. G. Paquette	To compute log-normal statistics on histogram data when the class intervals are variable or constant. For input it requires the top and bottom speed for each interval (number of intervals restricted by computer stowage) and the counts per interval. Its output includes printer plots of both arithmetic and logarithmic probability densities versus speeds, and associated statistical tables including the difference between the CDF of the data and the CDF of a normal distribution.
HISTG	D. W. Robinson Lt., USN Naval Postgraduate School, Monterey	To produce a one-page summary of a set of data by computing a set of basic statistics and printing a histogram. The capability exists to display a smoothed empirical density function plot on the histogram. The size of the data set is unlimited.
KOLMO	IBM	To test the difference between empirical and theoretical distributions using the Kolmogorov-Smirnov test. This program is part of IBM's SYSTEM/360 Scientific Subroutine Package Version III.

Table III. Summary Statistics of the Logarithmic-Speed Distribution of the Group of Current-Meter Time-Series Records at Designated Deviations from the Mean and the Results of a K-S Goodness-of-Fit Test of this Data to the Log-Normal.

Deviation From the Mean (Sigma Units)	Number of Series Used	Statistics for Values of OBS-PRED Cumulative Probability						K-S Goodness-of-Fit Test Results
		RANGE	RANGE	MEAN	Standard Deviation	Coefficient of Skewness	Coefficient of Kurtosis*	α
-3	29	.014	- .014	.002	.006	- .212	4.484	.000
-2	29	.037	- .028	.002	.015	- .250	3.065	.008
-1	29	.053	- .074	-.011	.029	.043	2.582	.000
-.5	29	.047	- .098	-.023	.036	- .509	2.815	.000
0	29	.080	- .122	-.022	.050	- .302	2.583	.000
.5	29	.100	- .093	-.007	.042	.748	3.883	.000
1	29	.099	- .037	.022	.041	.317	1.772	.000
2	28	.023	- .020	.008	.011	- .808	3.008	.000
3	23	.001	- .023	-.003	.007	- 1.851	5.088	.000

*Values Recorded as Computed without Subtracting 3.0

Table IV. Some Computations Used in the Analysis of Various Features of the Current-Meter Data Presented in Table III and Figure 6.

Dev. From the Mean (Sigma Units)	Number of Series Used (N)	Values for Analyzing the Coefficients of Skewness			Values for Analyzing the Significance of the Deviation of the Mean from zero.					
		Coefficient of Skewness	Standard Error (SE) *	2xSE	Mean	Standard Deviation	Standard Error (SE) of a Single Obs. +	t Value $t = \frac{MEAN}{SE}$	Degrees of Freedom $\nu = N-1$	Value from t-test α
-3	29	- .212	.434	.868	.002	.006	.0011	1.818	28	.08
-2	29	- .250	.434	.868	.002	.015	.0028	.714	28	.47
-1	29	.043	.434	.868	-.011	.029	.0054	2.037	28	.05
-.5	29	- .501	.434	.868	-.023	.036	.0067	3.433	28	.001
0	29	- .302	.434	.868	-.022	.050	.0093	2.366	28	.021
.5	29	.748	.434	.868	-.007	.042	.0078	.897	28	.37
1	29	.317	.434	.868	.022	.041	.0076	2.895	28	.01
2	28	- .808	.441	.882	.008	.011	.0021	3.810	27	.0007
3	23	- 1.851	.481	.962	-.003	.007	.0015	2.000	22	.05

* $SE = \left[\frac{6N(N-1)}{(N-2)(N+1)(N+3)}\right]^{\frac{1}{2}}$ + $SE = \sigma/N^{\frac{1}{2}}$

Table V. Coefficient of Skewness, Square of Coefficient of Skewness (β_1), and Coefficient of Kurtosis (β_2) for Logarithmic Current-Meter Data

Data Sample Number	Coefficient of Skewness	β_1	Coefficient of Kurtosis	Data Sample Number	Coefficient of Skewness	β_1	Coefficient of Kurtosis
NH-6 491/8	.66	.44	4.12	WF 1071	- .19	.04	9.91
NH-6 490/8	.39	.15	3.82	WF 1072	- 1.35	1.82	11.86
NH-15 455/5	-1.00	1.00	5.10	WF 1073	- 1.45	2.10	10.06
NH-15 456/5	- .56	.31	3.15	WF 1075	- .08	.01	10.60
NH-15 453/10	- .47	.22	3.07	WF 1076	.31	.10	7.84
NH-15 452/7	.45	.20	3.98	WF 1077	1.60	2.56	12.91
NH-15 455/10	- .40	.16	2.84	WF 1391	- .66	.44	5.61
NH-15 456/10	- .04	.00	2.55	WF 1392	- .83	.69	7.09
NH-15 D72/7	- .94	.88	3.79	WF 1393	- 1.06	1.12	4.69
NH-15 454/12	- .54	.29	3.46	WF 1395	1.39	1.93	8.78
NH-15 452/10	- .47	.22	2.88	WF 1398	.34	.12	4.76
SCARF 1	- .13	.02	6.93	WF 1401	- .12	.01	2.95
SCARF 2	- .24	.06	2.75	WF 1402	- .32	.10	2.46
SCARF 3	- .52	.27	3.01	WF 1403	- .47	.22	3.38
SCARF 4	- .33	.11	3.08	WF 1404	- .32	.10	3.53
SCARF 5	- .31	.10	3.65	WF 1531	.53	.28	5.70
SCARF 6	-1.30	1.69	11.17	WF 1533	- .78	.61	2.76
SCARF 7	- .22	.05	2.34	WF 1534	- .14	.02	3.17
WF 1011	- .54	.29	8.60	WF 1612	- .91	.83	6.00
WF 1012	.06	.00	2.88	—	—	—	—

Table VI. Statistical Summary of Drift-of-Ship Data.

MS AREA/ QUADRANT/ MONTH	NO. OF SPD CLASSES/ TOTAL COUNT	ARITHMETIC		LOGARITHMIC							
		$\overline{V}$ (cm/sec)	σ (cm/sec)	$\overline{\text{LOG V}}$	σ_L	COEFFICIENT OF SKEWNESS	COEFFICIENT OF KURTOSIS	AFTER ALTERATION		BEFORE ALTERATION	
								ΔPm	P	ΔPm	P
114-1-2	9/207	24.03	13.39	1.32	.24	-.843	4.768	-.0401	.6087	-.1088	.2802
114-1-3	12/1045	25.42	15.72	1.33	.25	-.316	3.016	-.0078	.8067	-.0351	.2775
114-1-5	11/1170	25.33	14.57	1.29	.26	-.536	3.929	-.0116	.6632	-.0954	.3419
114-1-6	10/283	25.03	15.64	1.32	.26	-.503	3.694	.0140	.9647	-.0647	.3039
114-1-7	8/345	23.08	13.60	1.29	.25	-.566	3.910	-.0152	.8377	-.1067	.3333
114-1-9	9/261	25.39	16.64	1.32	.27	-.432	3.609	-.0099	.8889	-.0653	.3142
114-1-10	8/255	26.70	14.51	1.36	.26	-.846	4.121	-.0676	.5137	-.0922	.5137
114-1-12	12/248	26.30	20.52	1.33	.27	-.325	4.206	.0218	.8185	-.1107	.2903
114-2-1	13/943	25.52	16.48	1.34	.25	-.241	3.002	.0120	.8271	-.0296	.2725
114-2-7	9/279	25.08	15.27	1.32	.27	-.638	3.862	-.0284	.6022	-.0997	.3226
114-4-4	10/188	32.97	21.62	1.42	.31	-.723	3.868	-.0394	.4468	-.0852	.4468
115-1-1	11/957	26.07	16.94	1.34	.25	-.216	2.928	.0076	.8109	-.0239	.2696
115-1-7	11/354	29.79	21.21	1.38	.29	-.471	3.716	-.0114	.7175	N/A	N/A
115-1-10	9/267	27.32	16.59	1.36	.26	-.642	4.171	-.0152	.7566	-.1091	.2368
115-1-11	11/891	24.71	17.11	1.31	.27	-.423	3.738	.0137	.3468	-.0888	.3468
115-1-12	10/247	24.15	14.60	1.31	.25	-.587	4.046	.0114	.9960	-.0995	.3036
115-2-5	10/1155	24.46	15.52	1.31	.27	-.658	4.266	-.0324	.6139	-.1163	.3229

Table VI. Continued.

MS AREA/ QUADRANT/ MONTH	NO. OF SPD. CLASSES/ TOTAL COUNT	ARITHMETIC		LOGARITHMIC							
		$\overline{V}$ (cm/sec)	σ (cm/sec)	$\overline{\text{LOG V}}$	σ_L	COEFFICIENT OF SKEWNESS	COEFFICIENT OF KURTOSIS	AFTER ALTERATION		BEFORE ALTERATION	
								ΔPm	P	ΔPm	P
115-2-11	12/861	25.27	18.73	1.31	.29	- .390	3.693	.0254	.3624	-.0753	.3624
115-2-12	12/248	26.27	19.31	1.33	.28	- .384	3.833	.0224	.9597	-.0916	.3065
115-3-1	12/623	44.65	29.90	1.56	.29	- .385	2.879	-.0294	.6164	-.0291	.6164
115-3-7	11/286	42.07	30.09	1.51	.33	- .592	3.669	-.0239	.8322	-.0934	.3566
115-3-9	12/236	45.75	33.63	1.54	.34	- .513	2.997	-.0354	.6102	-.0760	.3305
115-4-1	13/832	39.42	29.36	1.49	.30	- .212	3.123	.0327	.6154	-.0375	.1406
115-4-7	14/439	43.58	30.86	1.54	.31	- .731	4.248	-.0420	.3030	-.1074	.3030
115-4-11	12/1379	43.00	31.23	1.52	.33	- .571	3.500	-.0306	.6410	-.0715	.3481
116-1-7	14/1376	32.66	26.05	1.41	.31	- .516	4.192	-.0136	.4913	-.1200	.2362
116-1-8	13/4724	33.48	23.50	1.43	.28	- .352	3.426	.0098	.8815	-.0412	.1854
116-3-1	14/4855	39.91	31.06	1.49	.31	- .338	3.500	.0141	.8140	-.0560	.1617
116-3-7	14/2540	45.55	39.46	1.52	.36	- .452	3.629	-.0113	.3736	-.0956	.1913
149-1-1	10/487	32.23	20.01	1.43	.26	- .352	2.744	-.0182	.6653	-.0186	.1807
149-1-7	10/282	33.35	20.51	1.44	.29	- .762	4.033	-.0398	.4220	-.0832	.4220
149-3-1	10/344	30.41	21.54	1.39	.28	- .106	2.490	-.0222	.8140	-.0182	.8140
149-3-2	8/58	27.91	18.17	1.35	.30	- .742	4.036	-.0632	.6897	-.0645	.3103
149-3-3	11/410	26.64	17.65	1.35	.25	- .300	3.346	.0181	.8098	-.0540	.2585

Table VI. Continued.

MS AREA/ QUADRANT/ MONTH	NO. OF SPD. CLASSES/ TOTAL COUNT	ARITHMETIC		LOGARITHMIC							
		$\overline{V}$ (cm/sec)	σ (cm/sec)	$\overline{\text{LOG V}}$	σ_L	COEFFICIENT OF SKEWNESS	COEFFICIENT OF KURTOSIS	AFTER ALTERATION		BEFORE ALTERATION	
								ΔPm	P	ΔPm	P
149-3-5	10/466	27.36	16.80	1.36	.27	- .573	3.622	-.0208	.5558	-.0769	.2651
149-3-6	8/91	28.43	15.90	1.38	.26	-1.028	5.358	-.0678	.4725	-.0930	.4725
149-3-7	14/127	30.51	25.36	1.38	.30	- .364	3.814	-.0191	.5197	-.0701	.2677
149-3-9	10/121	28.98	20.98	1.36	.31	- .372	3.021	.0487	.3388	-.0735	.5289
149-3-10	8/81	27.40	16.63	1.35	.29	- .752	3.956	-.0489	.5309	-.0875	.2963
149-3-11	14/549	28.14	20.17	1.36	.28	- .533	4.003	-.0333	.5357	-.0988	.2750
149-3-12	9/46	23.79	16.28	1.30	.25	- .048	2.663	.0462	.7174	-.1052	.3261
149-4-1	10/236	30.90	17.56	1.42	.24	- .459	3.015	.0163	.9788	-.0240	.3489
149-4-3	10/177	27.08	15.74	1.37	.24	- .385	2.678	-.0313	.7514	.0449	.9153
149-4-7	10/62	26.68	18.06	1.35	.25	- .071	2.896	.0352	.6290	-.0931	.2419
150-1-7	12/302	36.17	26.28	1.45	.31	- .480	3.486	-.0282	.7252	-.0622	.2086
150-2-1	13/865	35.47	24.60	1.46	.28	- .246	2.725	.0098	.9988	.0116	.9988
150-2-2	10/111	30.74	19.49	1.41	.26	- .208	2.803	.0524	.5495	-.0852	.1712
150-2-7	14/348	37.28	28.14	1.46	.31	- .339	3.072	-.0183	.7241	-.0488	.4138
151-1-1	14/848	30.44	21.43	1.40	.27	- .242	3.075	.0180	.9233	.0276	.9233
151-1-10	14/343	29.50	23.17	1.37	.30	- .372	3.675	.0103	.2828	-.0802	.2828
—	—	—	—	—	—	—	—	—	—	—	—

Table VII. Summary Statistics of the Logarithmic-Speed Distribution of the Group of Drift-of-Ship Current Records at Designated Deviations from the Mean, and the Results of a K-S Goodness-of-Fit Test of this Data to the Log-Normal.

DEVIATION FROM THE MEAN (SIGMA UNITS)	NUMBER OF DATA RECORDS USED	RANGE		MEAN	STANDARD DEVIATION	COEFFICIENT OF SKEWNESS	COEFFICIENT OF KURTOSIS *	K-S GOODNESS-OF-FIT TEST RESULTS α
- 2	50	.007	- .009	- .001	.003	.220	3.470	.000
- 1	50	.032	- .018	.003	.010	.691	4.722	.006
-.5	50	.042	- .034	- .001	.012	.421	6.663	.008
0	50	.051	- .062	- .009	.019	.115	5.200	.000
.5	50	.040	- .049	- .010	.016	.178	4.015	.000
1	50	.028	- .024	- .002	.011	.224	2.812	.042
2	49	.021	- .010	.007	.007	.160	2.461	.000
3	21	.001	- .007	- .001	.002	-1.858	5.780	.000

*Values Recorded as Computed without Subtracting 3.0

Table VIII. Some Computations Used in the Analysis of Various Features of the Drift-of-Ship Data Presented in Table VII and Figure 14.

DEV. FROM THE MEAN (SIGMA UNITS)	NUMBER OF RECORDS USED (N)	VALUES FOR ANALYZING THE COEFFICIENTS OF SKEWNESS			VALUES FOR ANALYZING THE SIGNIFICANCE OF THE DEVIATION OF THE MEAN FROM ZERO					
		COEFFICIENT OF SKEWNESS	STANDARD ERROR * (SE)	2xSE	MEAN	STANDARD DEVIATION	STANDARD ERROR(SE) OF A SINGLE OBSERVATION +	t value $t = \frac{MEAN}{SE}$	DEGREES OF FREEDOM $\nu = N-1$	VALUE FROM t-test α
-2	50	.220	.3366	.6732	-.001	.003	.00042	2.381	49	.015
-1	50	.691	.3366	.6732	.003	.010	.00141	2.128	49	.032
-.5	50	.421	.3366	.6732	-.001	.012	.00170	.588	49	.540
0	50	.115	.3366	.6732	-.009	.019	.00269	3.346	49	.005
.5	50	.178	.3366	.6732	-.010	.016	.00226	4.425	49	.000
1	50	.224	.3366	.6732	-.002	.011	.00156	1.282	49	.200
2	49	.160	.3398	.6796	.007	.007	.00100	7.000	48	.000
3	21	-1.858	.5012	1.0024	-.001	.002	.00044	2.273	20	.034

$$* \quad SE = \left[\frac{6N(N-1)}{(N-2)(N+1)(N+3)} \right]^{\frac{1}{2}}$$

$$+ \quad SE = \frac{\sigma}{N^{\frac{1}{2}}}$$

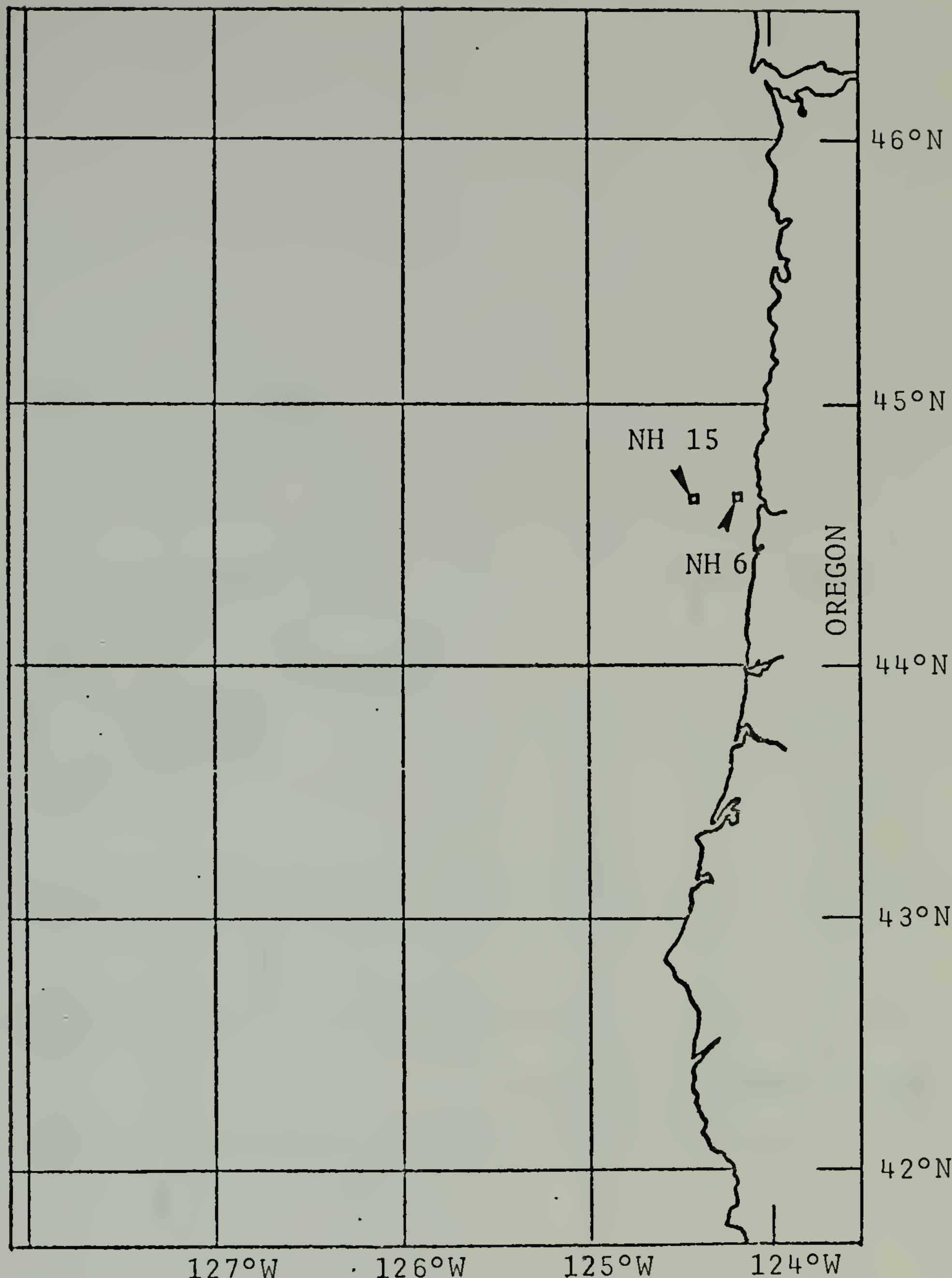


Figure 1. Location of Oregon State University Coastal Upwelling Experiment Current Meters

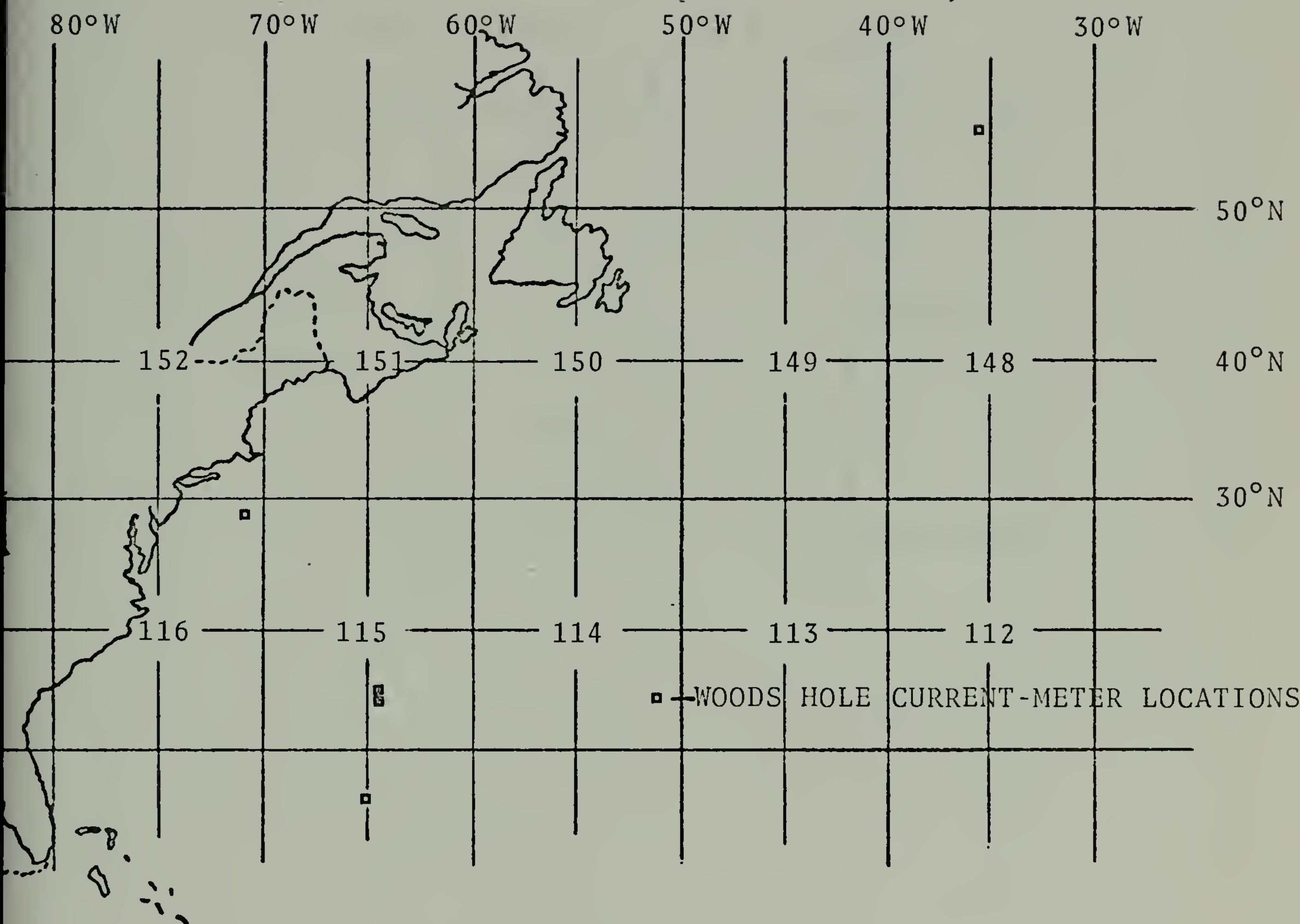


Figure 2. Marsden Square Grid Off the East Coast

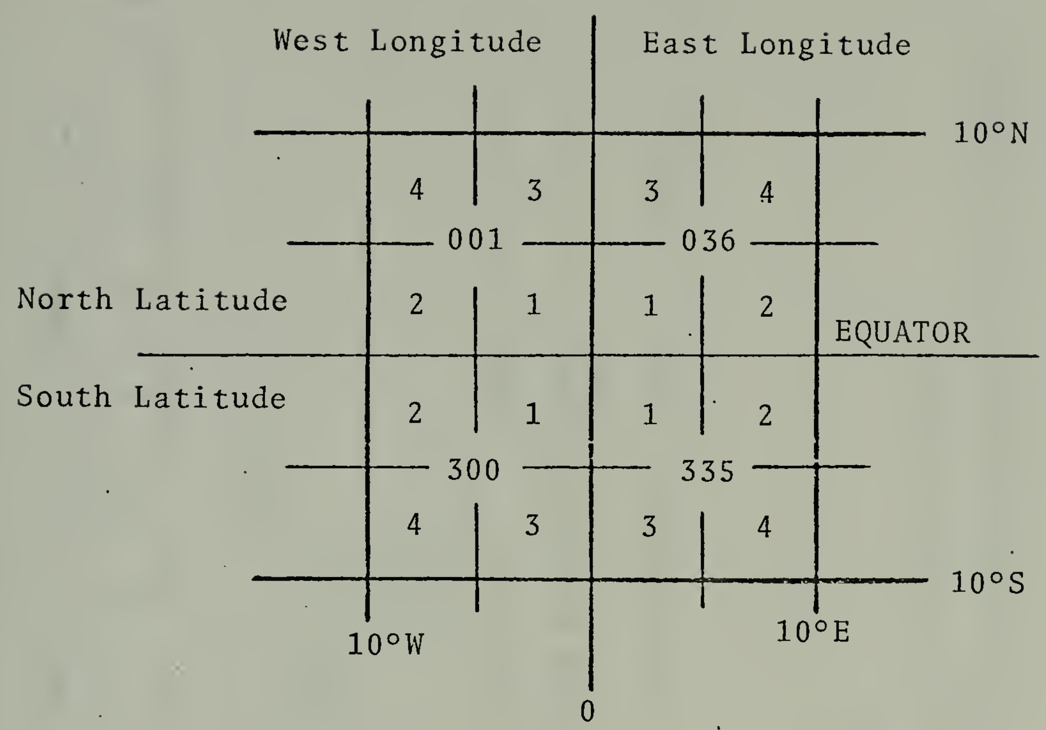


Figure 3. Marsden Square Grid System.

SURFACE CURRENT SUMMARY H 1-9

S SC	MO	DIR	C CALM	L 0.1	A 0.3	S 0.5	S 0.7	0.9	I 1.1	N 1.4	T 1.7	E 2.0	R 2.5	V 3.0	A 3.5	L 4.0	:S	SUM OBS	PCT OBS
1	10	CALM	0019															00019	007.1
		N		0005	0006	0004	0005			0001								00021	007.9
		NE		0006	0013	0001	0005	0004										00029	010.9
		E		0008	0014	0005	0005	0003	0002									00037	013.9
		SE		0006	0009	0007	0002	0002	0002									00028	010.5
		S		0004	0010	0003	0003		0003									00023	003.6
		SW		0004	0006	0010	0004		0001									00025	009.4
		W		0005	0022	0017	0005	0002	0002	0003								00056	021.1
		NW		0006	0010	0001	0006	0002	0001		0002							00028	010.5
SUM OBS		=	0019	0044	0090	0048	0035	0013	0011	0004	0002							00266	
PCT OBS		=	07.1	16.5	33.8	18.0	13.2	04.9	04.1	01.5	00.8								100

RC(DIR) = 272 RC(SPEED) = 0.05 AVG. SPEED = 0.47 V(N) = +0.00 V(E) = -0.05

LEGEND

RC(DIR) - ARCTAN [V(E)/V(N)]

RC(SPEED) - $[V(N)^2 + V(E)^2]^{\frac{1}{2}}$

AVG SPEED - ARITHMETIC AVERAGE OF CURRENT SPEEDS

V(N),V(E) - AVERAGE NORTHERN AND EASTERN COMPONENTS OF RESULTANT CURRENT SPEED

Figure 4. NODC Computer-Generated Printout of DOS Data for MS 115 Quadrant 1 month 10 (October)

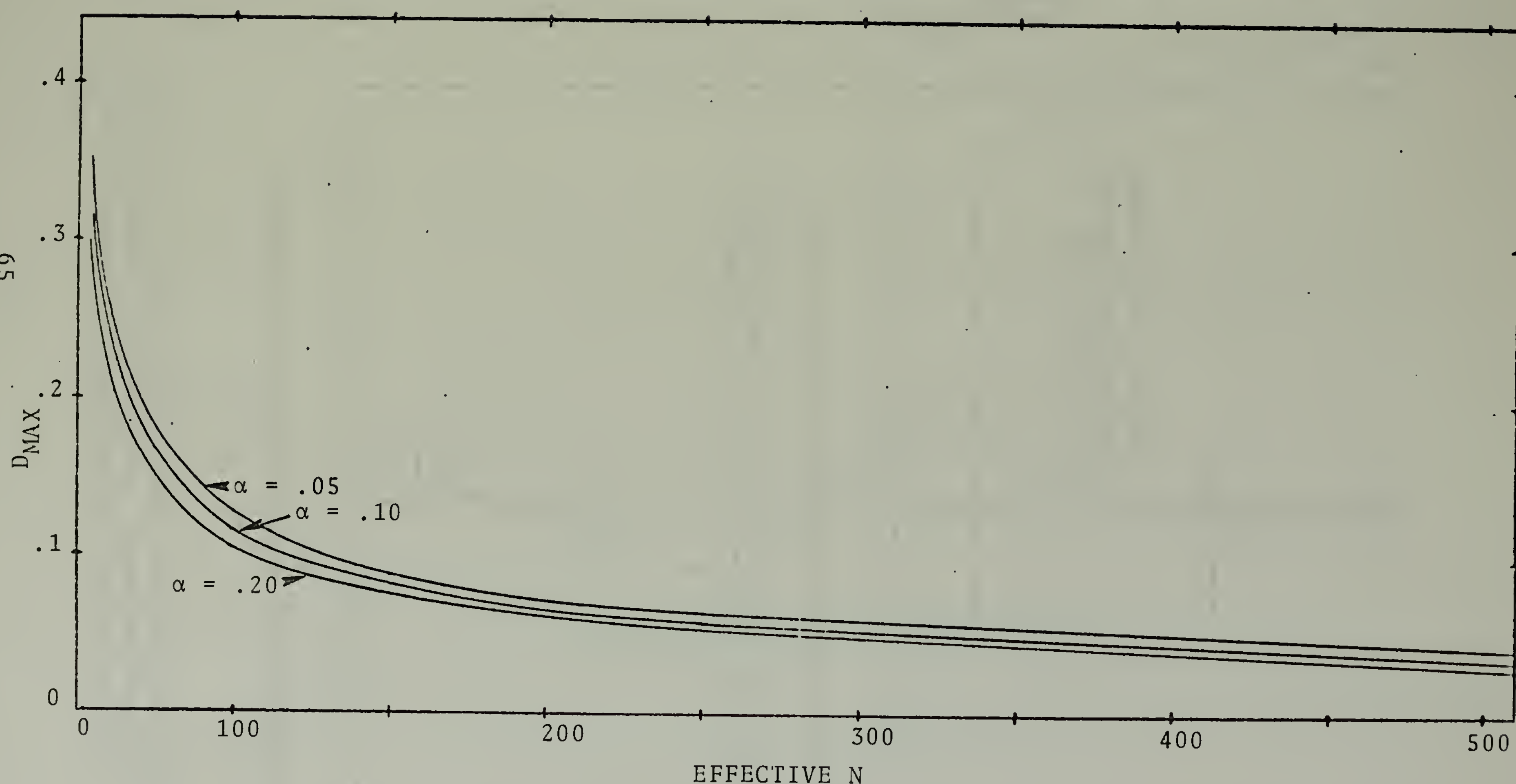


Figure 5. The Kolmogorov-Smirnov Statistic When the Mean and Standard Deviation of a Parent Distribution are Estimated from the Data

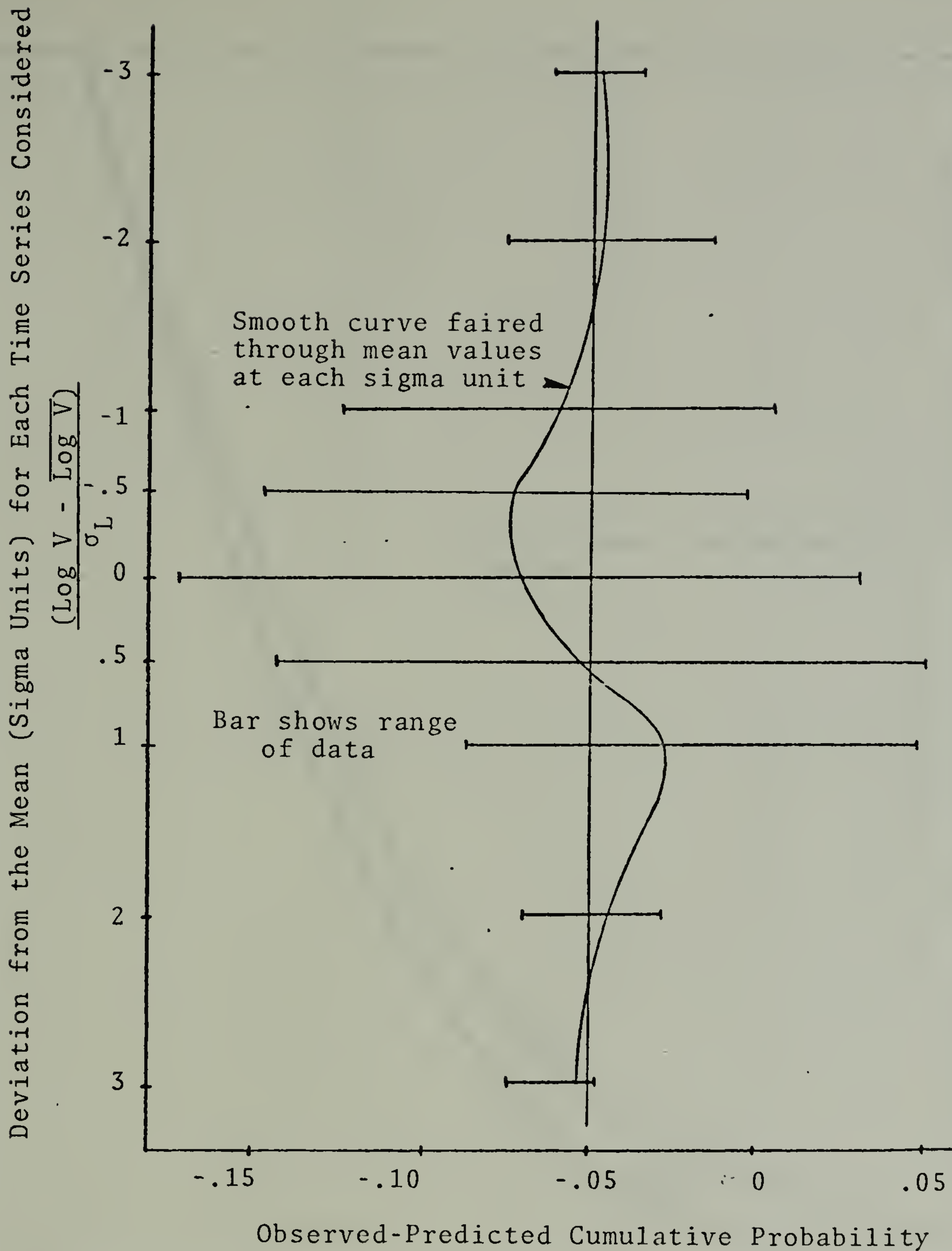


Figure 6. Deviation of Logarithmic-Speed Distribution from the Log-Normal Distribution. Mean and Range of 23 to 29 Moored Current-Meter Time-Series Data Sets.

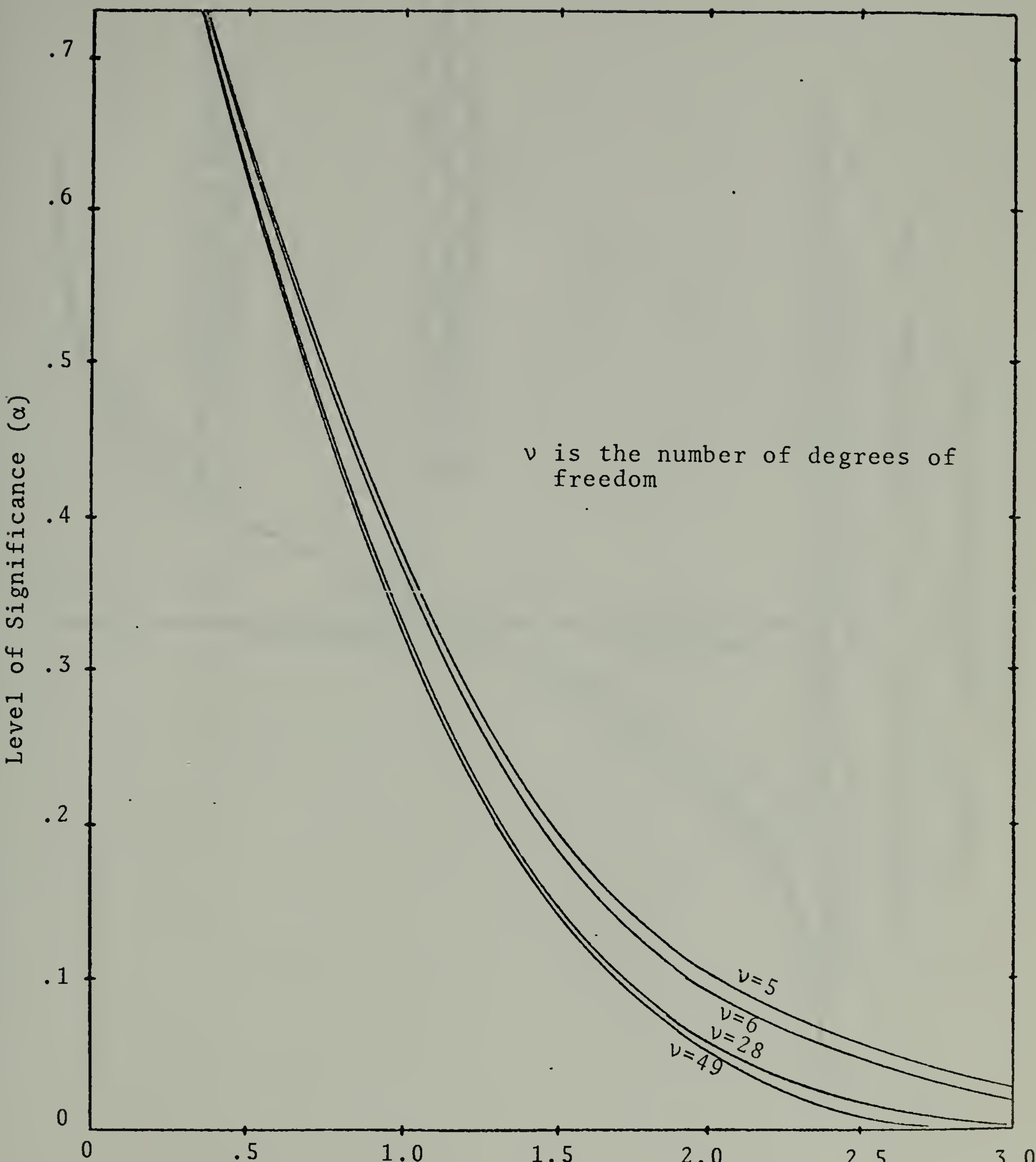


Figure 7. The "Student" t Statistic.

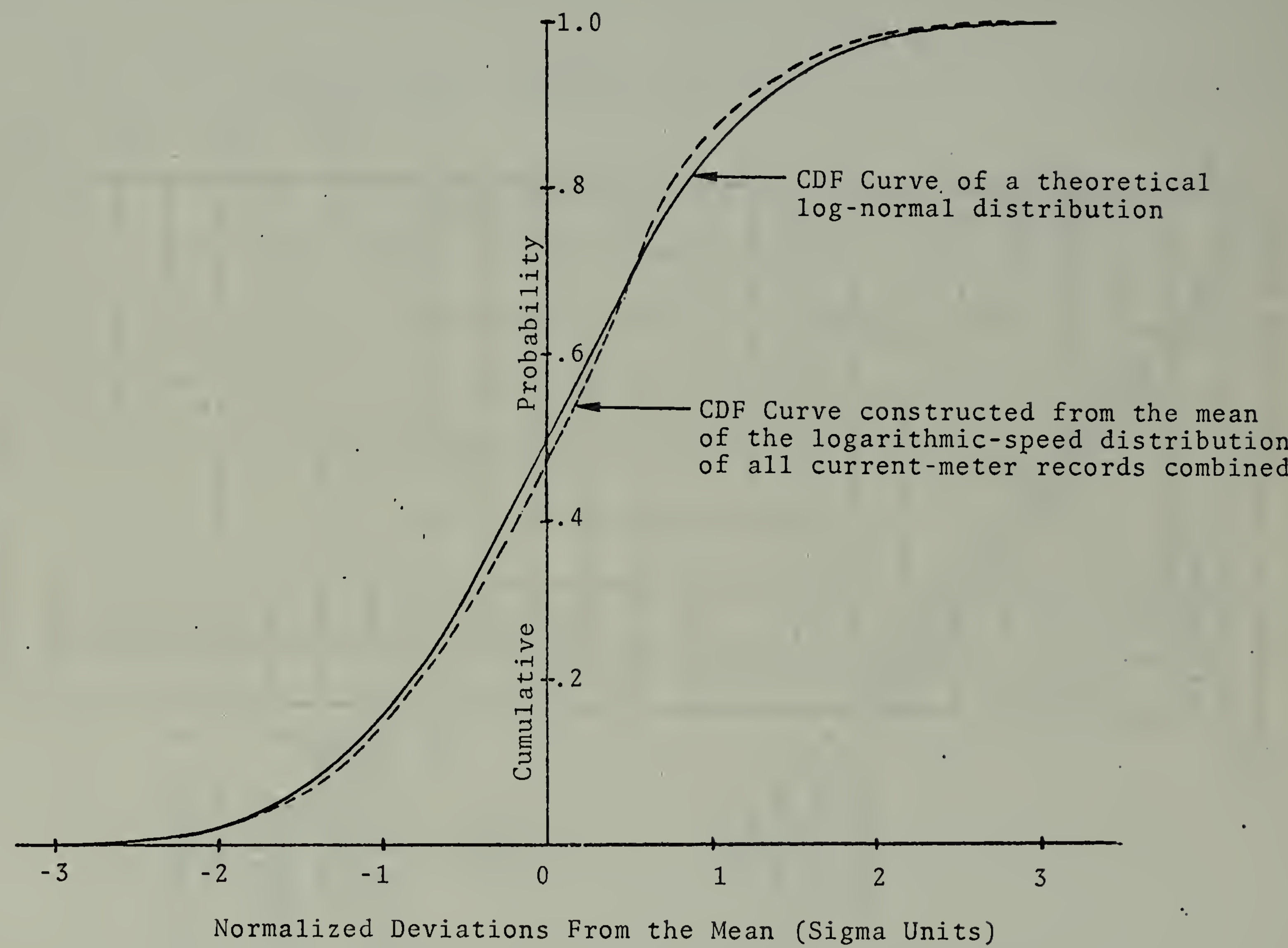


Figure 8. Comparison of CDF Curves.

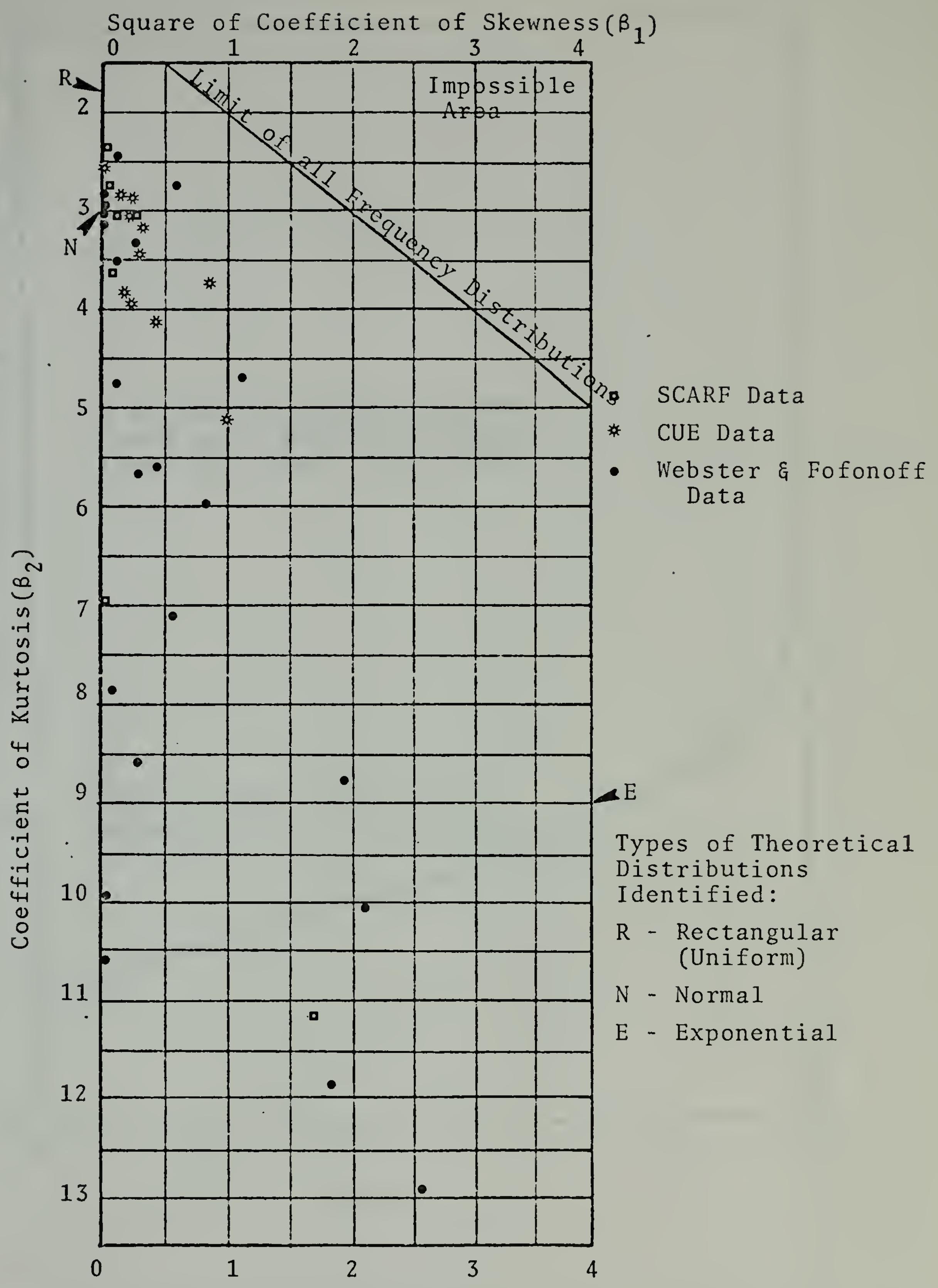


Figure 9. Pearson Diagram of β_1, β_2 Values for Logarithmic Current-Meter Data.

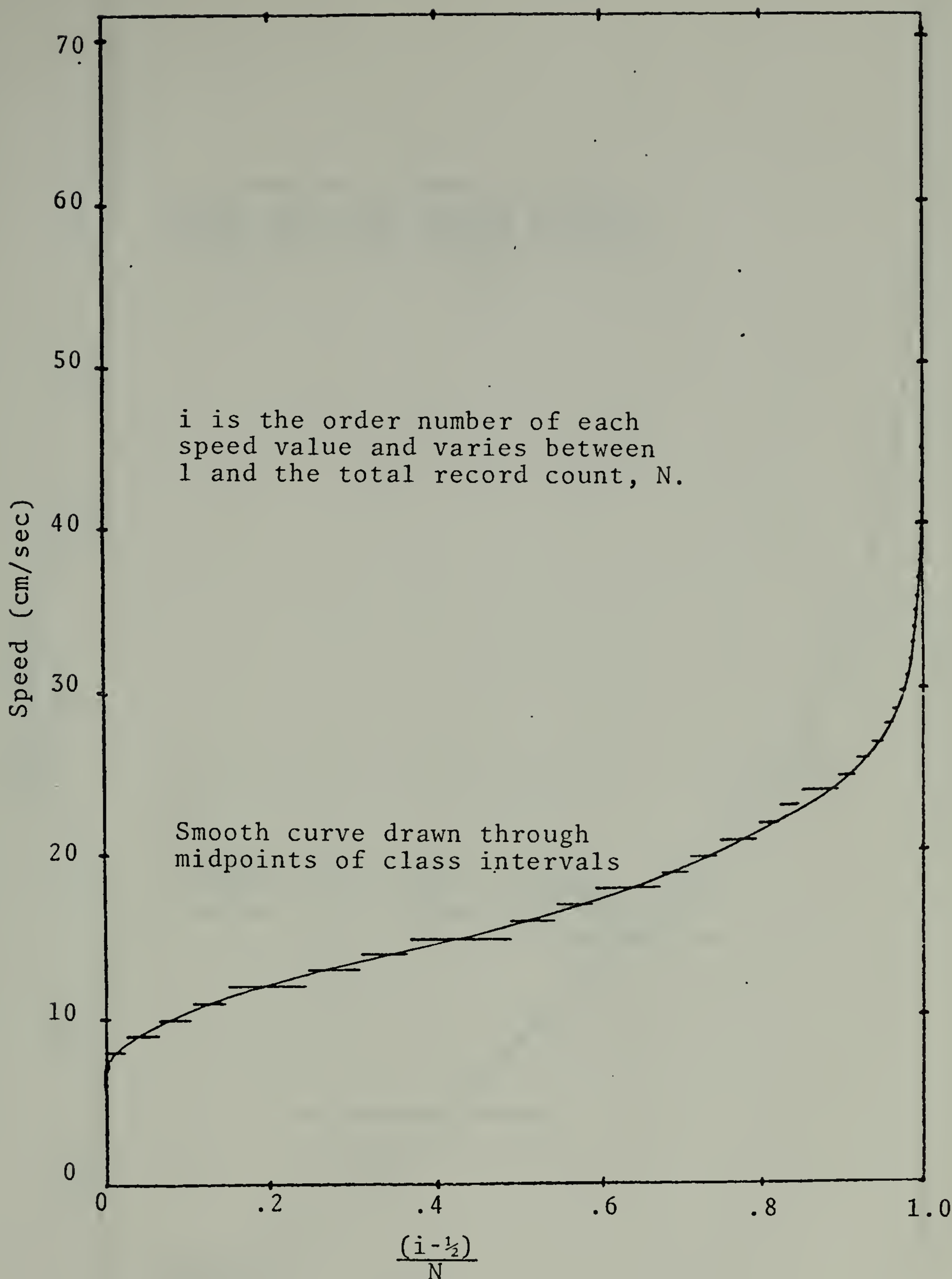


Figure 10. Empirical Cumulative Distribution Function (e.c.d.f.) for WF 1012.

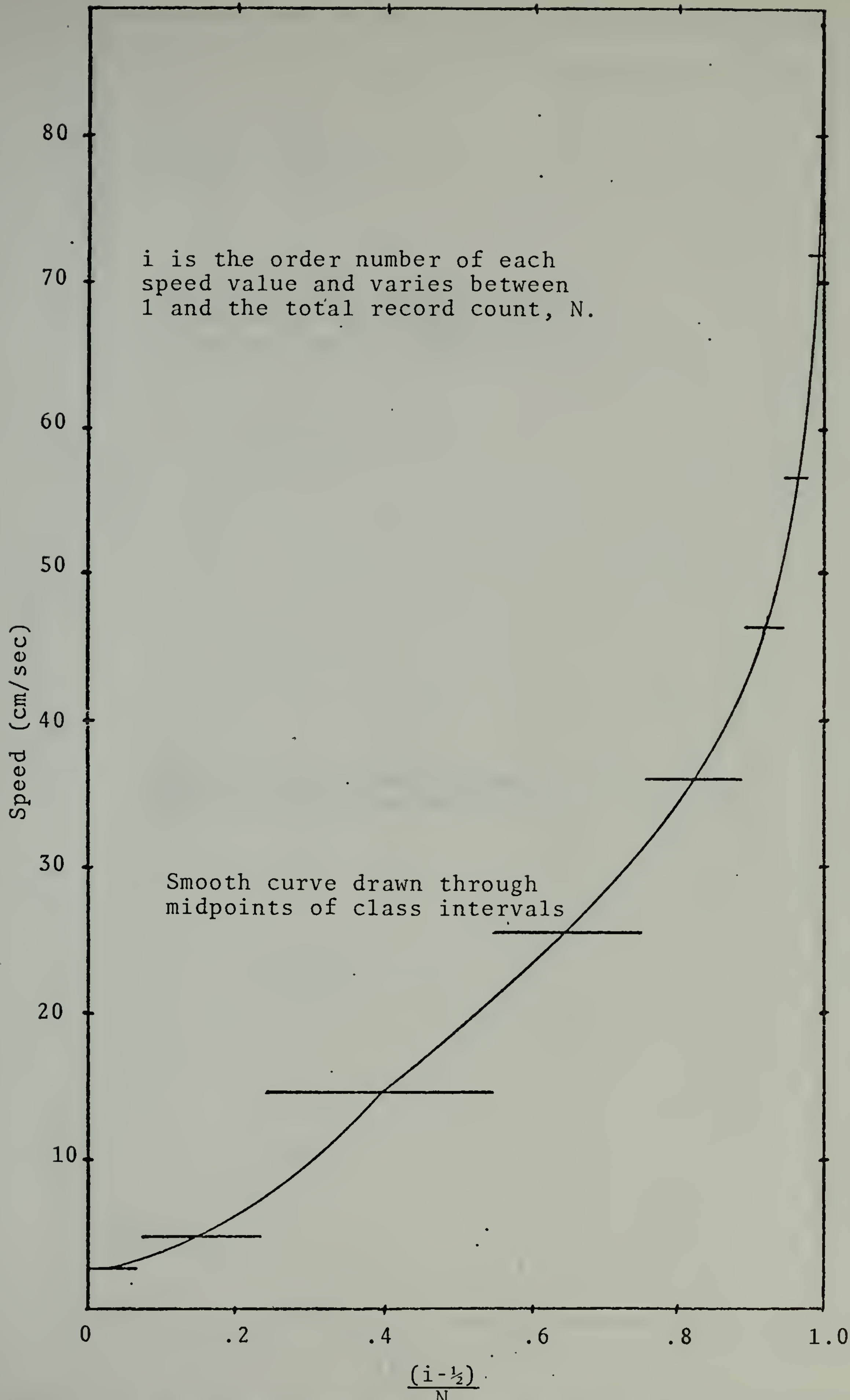


Figure 11. Empirical Cumulative Distribution Function (e.c.d.f.)

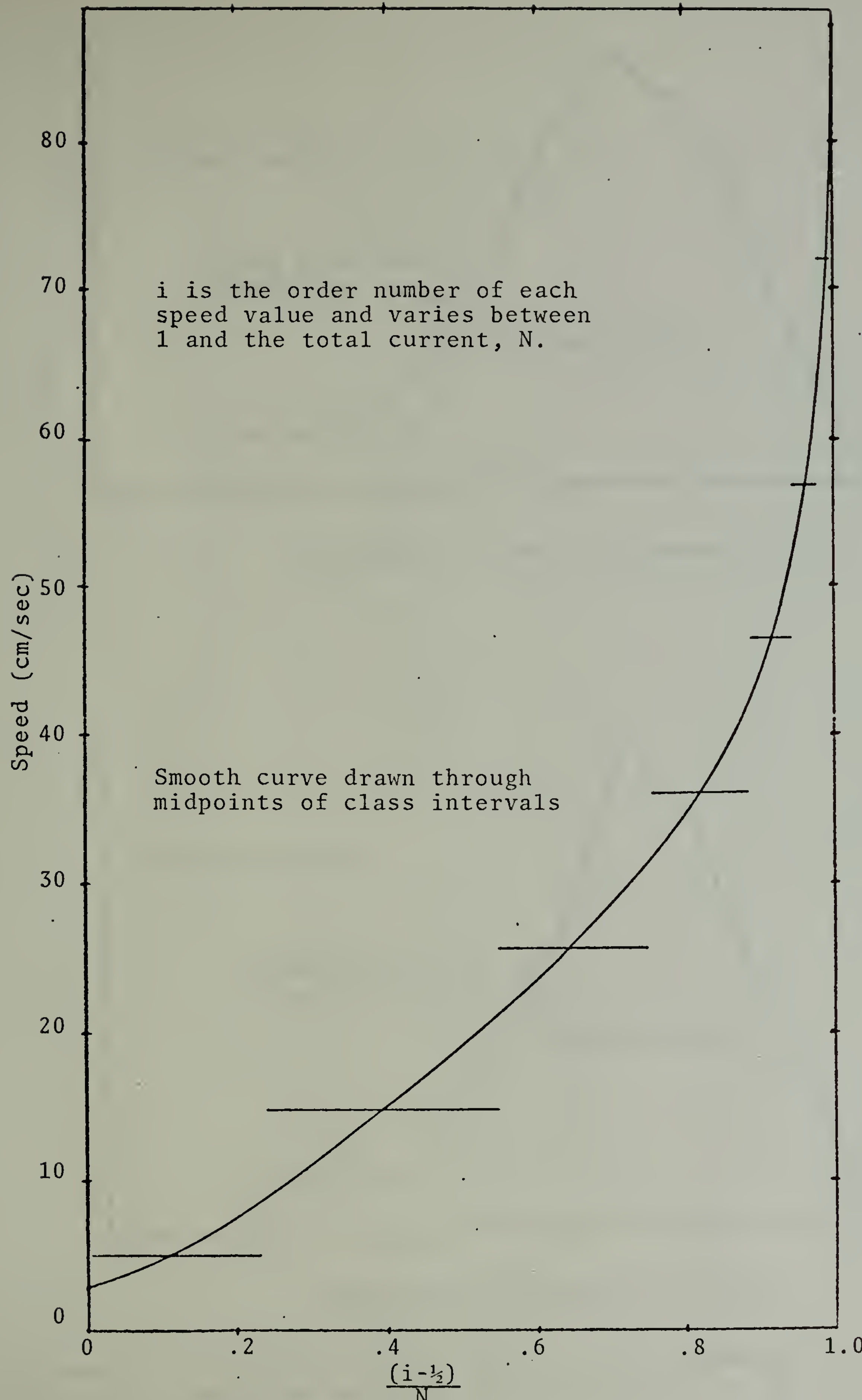


Figure 12. Empirical Cumulative Distribution Function (e.c.d.f.) for MS 115-1-10 After Nine-Tenths Alteration.

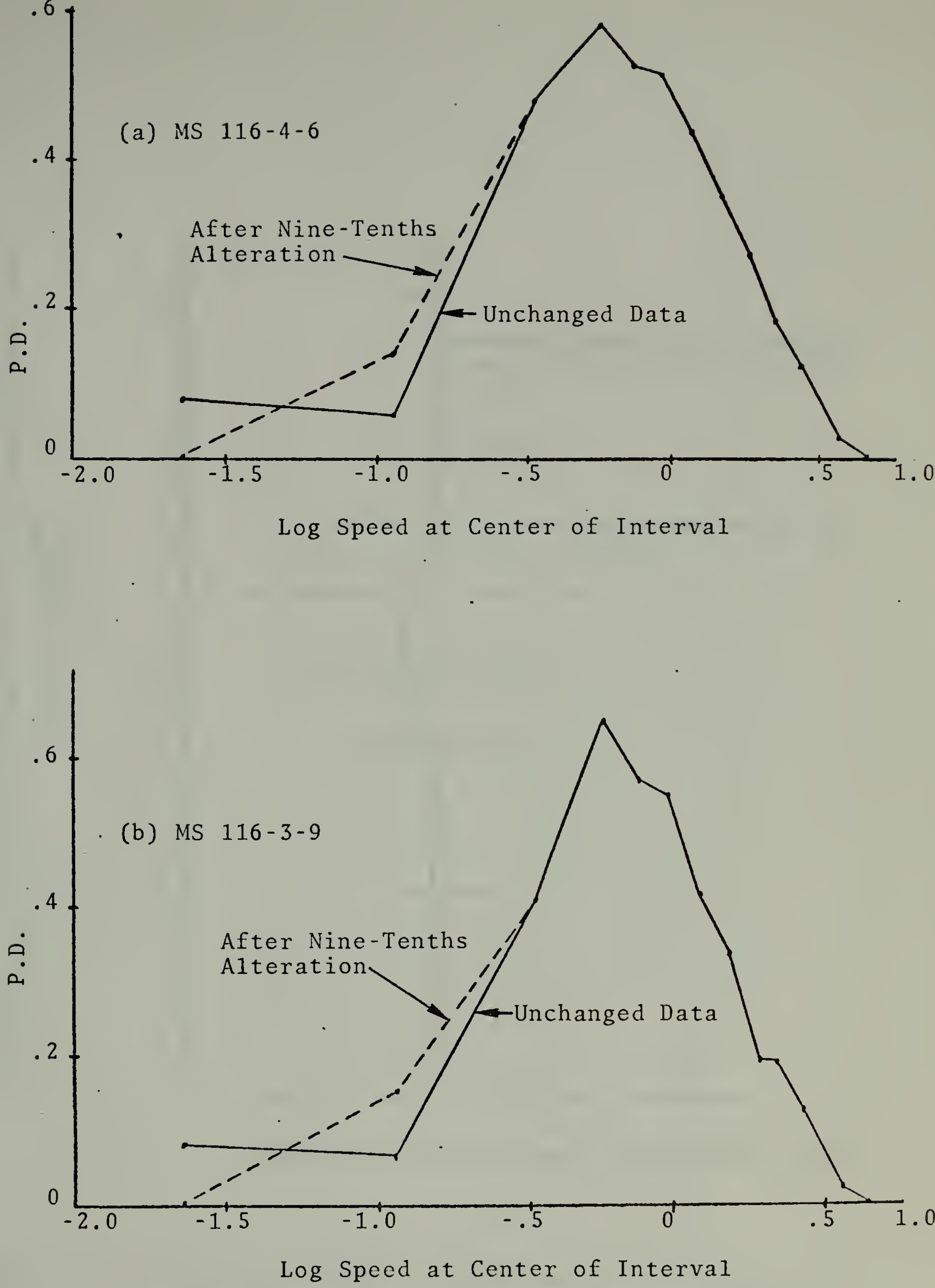
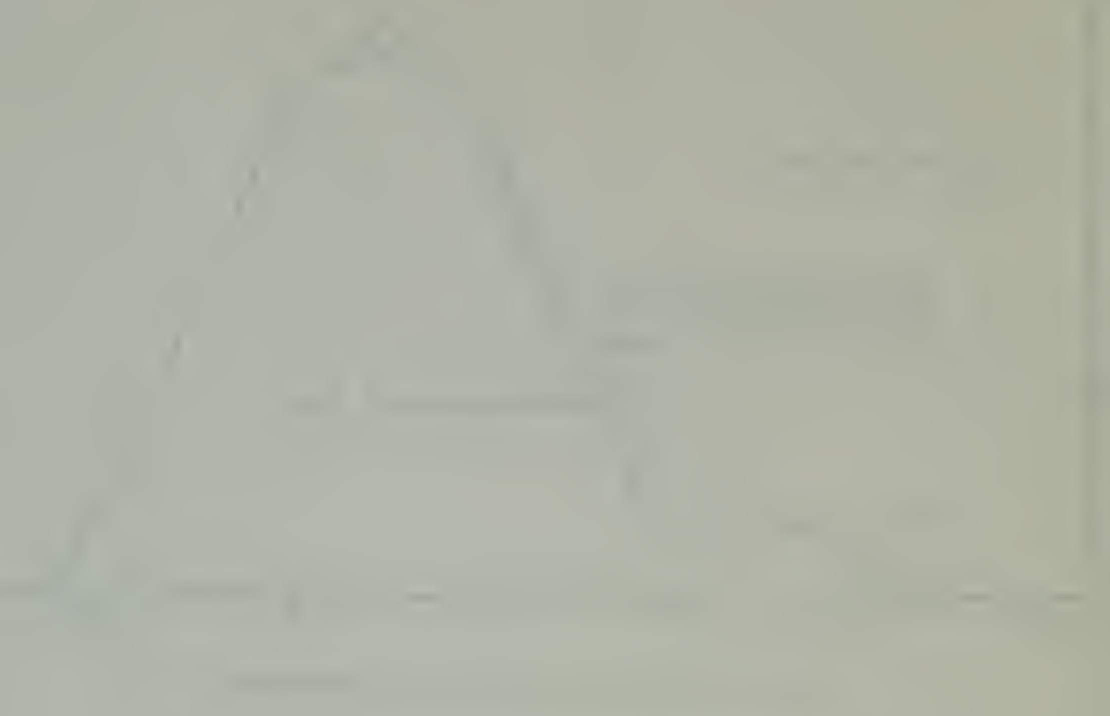
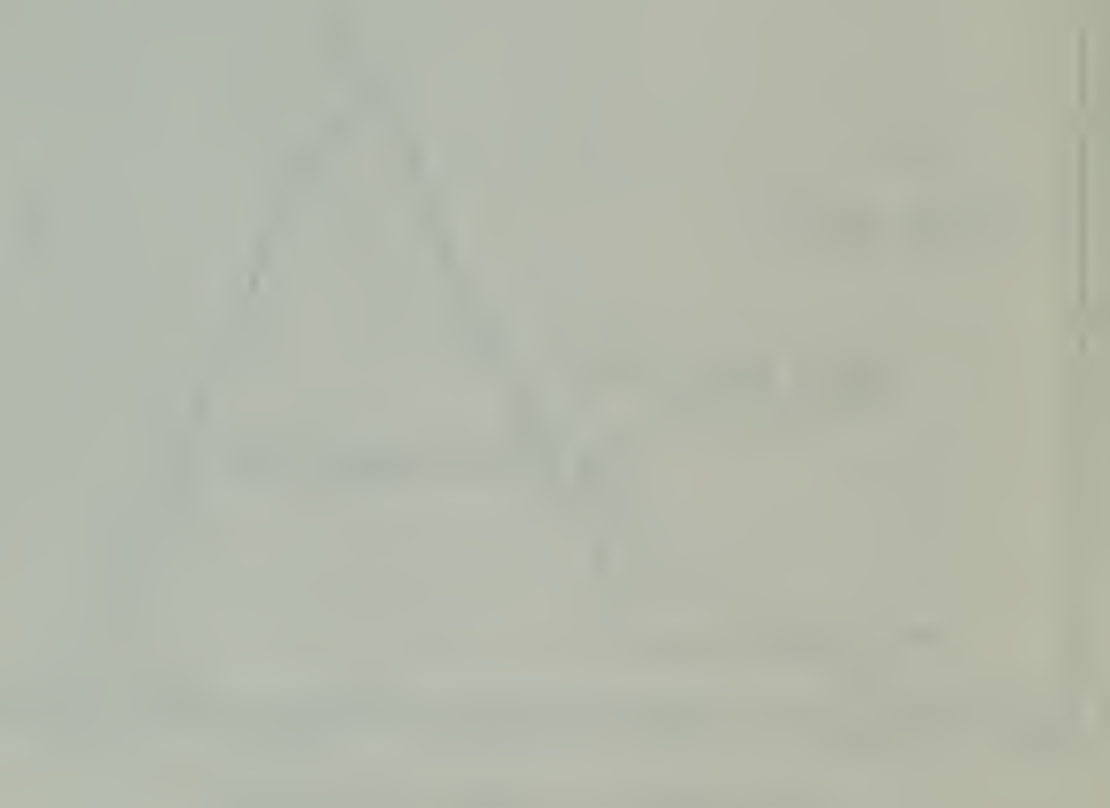


Figure 13. Probability Density Plots of the Logarithmic-Speed Distribution for Drift-of-Ship Data (a) MS 116-4-6 and (b) MS 116-3-9.





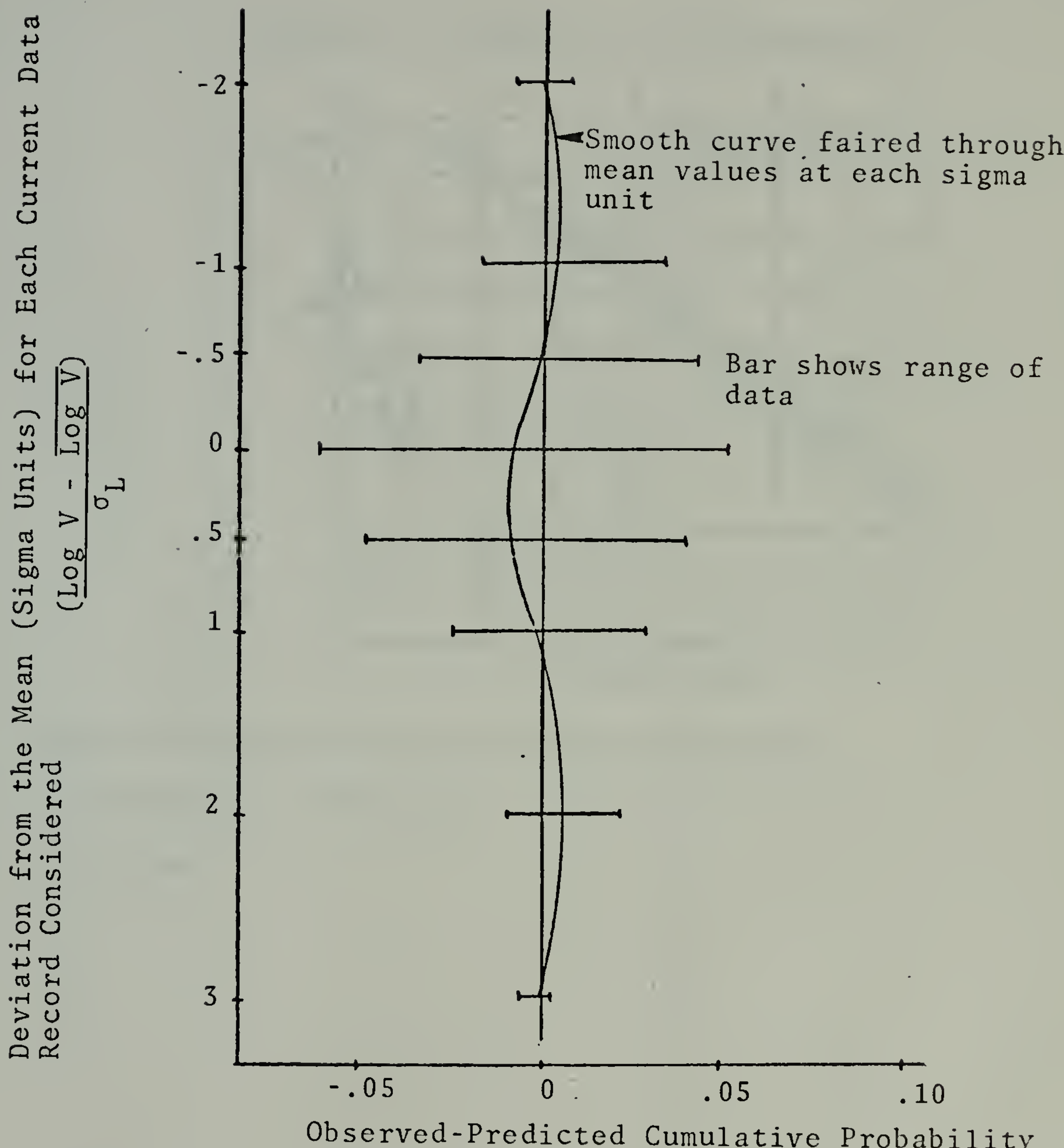
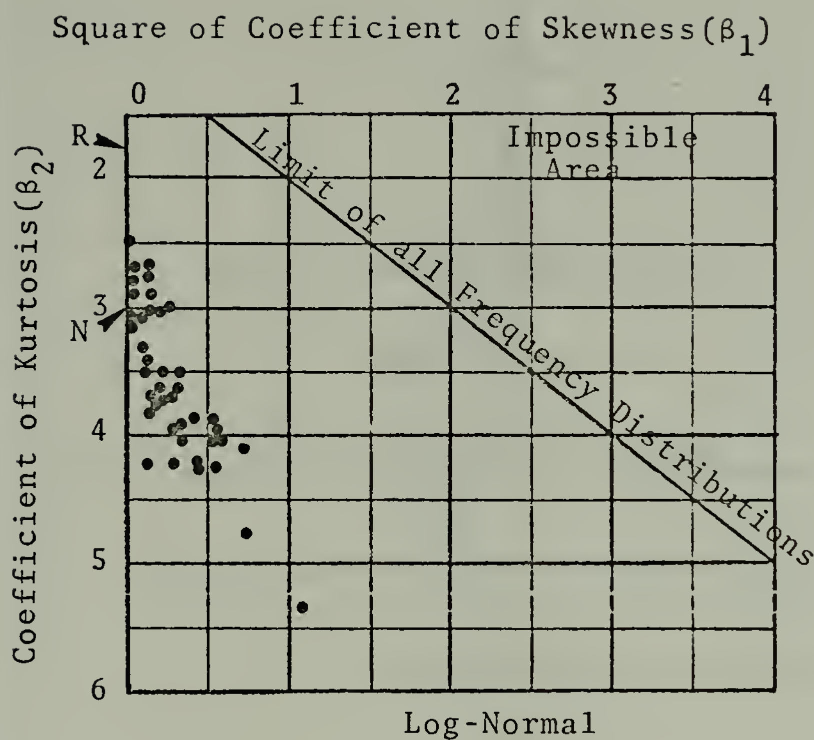


Figure 14. Deviation of Logarithmic-Speed Distribution from the Log-Normal Distribution. Mean and Range of 50 Altered Drift-of-Ship Current-Data Records.



Types of Theoretical Distributions Identified:

R - Rectangular (Uniform)

N - Normal

Figure 15. Pearson Diagram of the β_1, β_2 Values for Logarithmic Altered Drift-of-Ship Data.

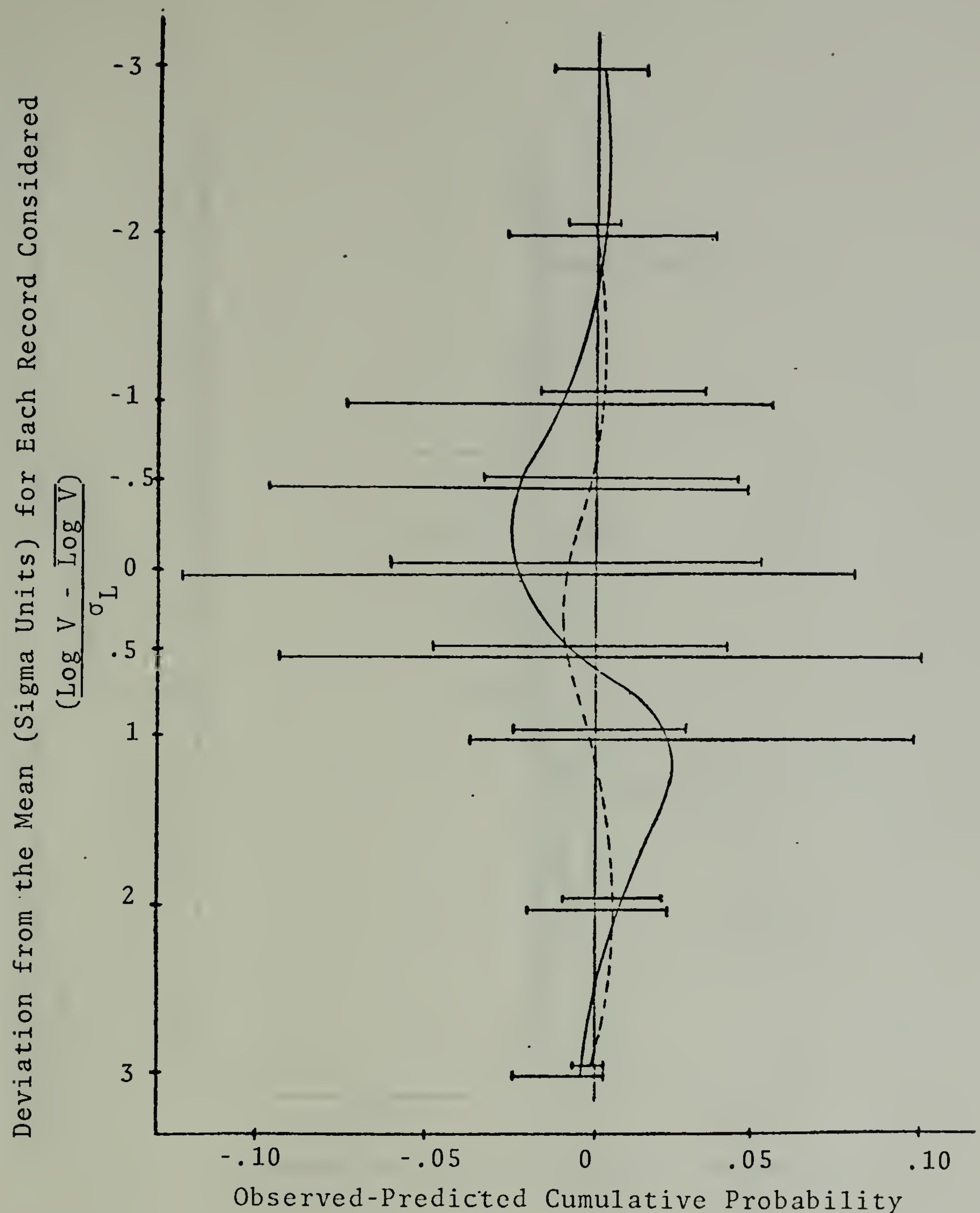


Figure 16. Comparison of Figure 6 (Solid Line, Lower Bars) and Figure 14 (Dashed Line, Upper Bars).

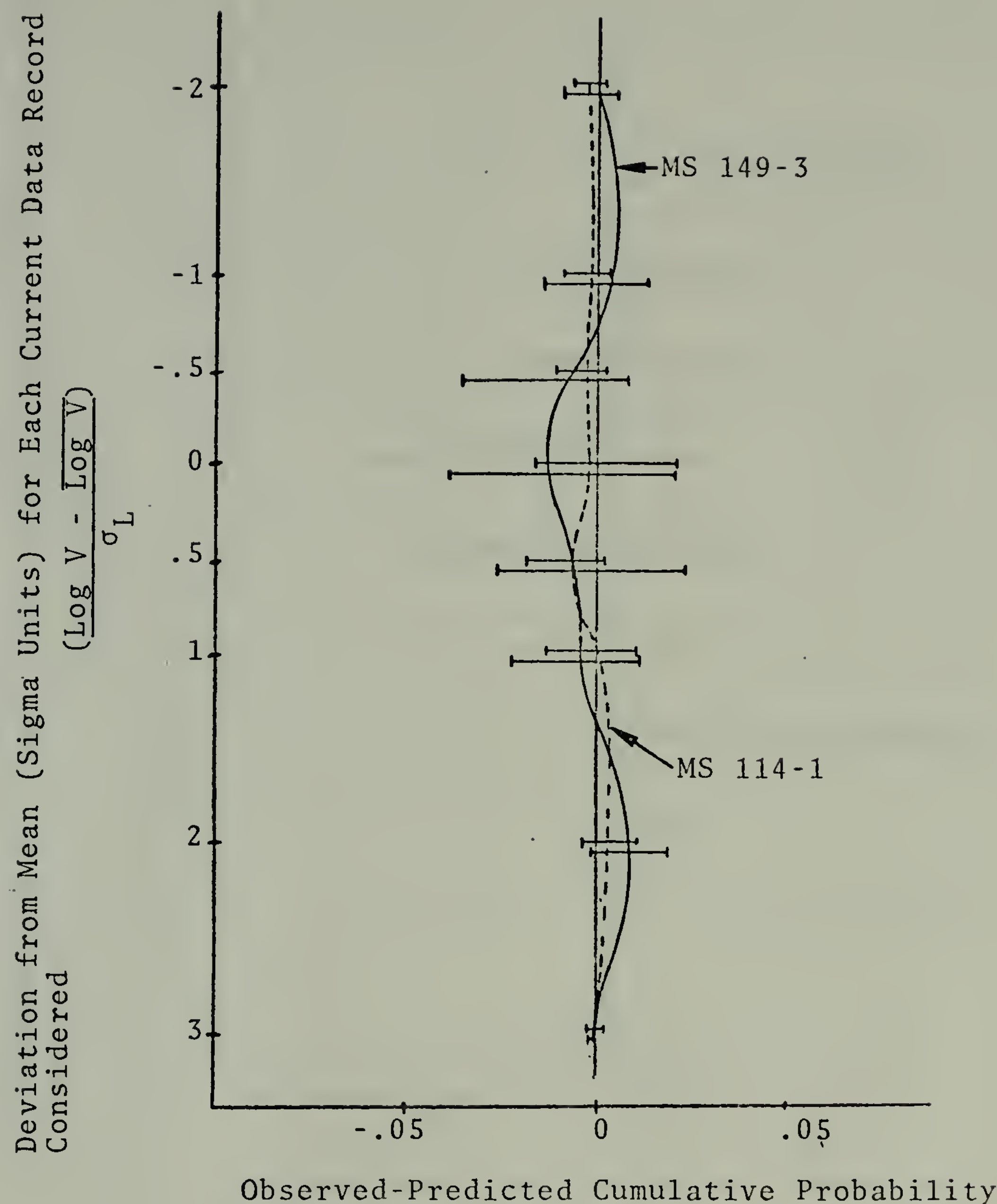


Figure 17. Deviation of Logarithmic-Speed Distribution from the Log-Normal Distribution. Comparison of the Mean and Range for Two Separate Areas; MS 114-1 (Dashed Line, Upper Bars) and MS 149-3 (Solid Line, Lower Bars).

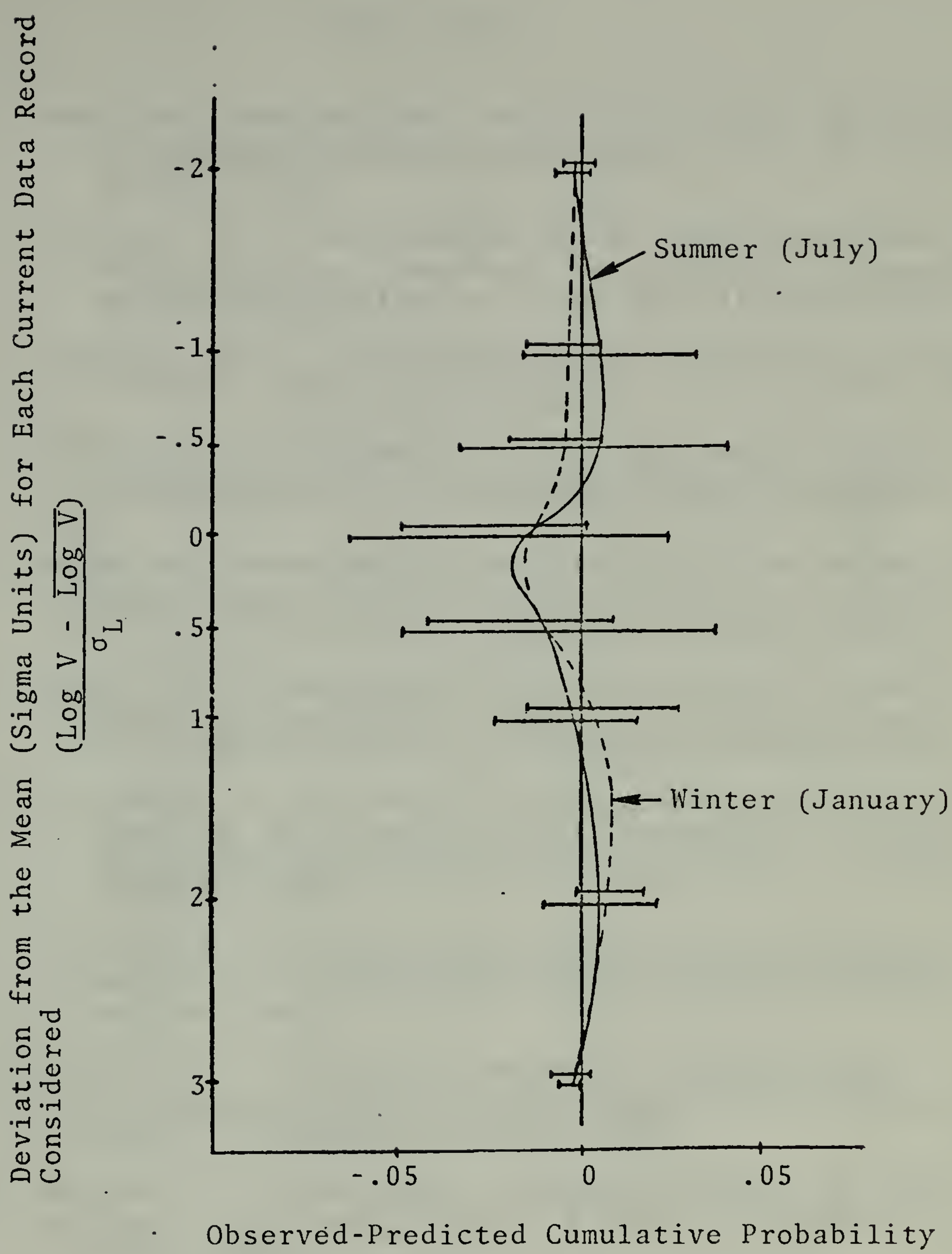


Figure 18. Deviation of Logarithmic-Speed Distribution from the Log-Normal Distribution. Comparison of the Mean and Range for Two Separate Seasons; January (Dashed Line, Upper Bars) and July (Solid Line, Lower Bars).

BIBLIOGRAPHY

1. Woods Hole Oceanographic Institution Report 64-55, Processing Moored Current Meter Data, by F. Webster, December 1964.

2. Belyayev, V. S. and Ozmidov, R. V., "Distributions of Velocity Vector Components in the Ocean," Atmospheric and Oceanic Physics, v. 7, p. 528-533, May 1971.

3. Paquette, R. G., "Some Statistical Properties of Ocean Currents," Ocean Engineering, v. 2, p. 95-114, 1972.

4. Woods Hole Oceanographic Institution Report 65-44, A Compilation of Moored Current Meter Observations, by F. Webster and N. P. Fofonoff, August 1965.

5. Woods Hole Oceanographic Institution Report 66-60, A Compilation of Moored Current Meter Observations, by F. Webster and N. P. Fofonoff, November 1966.

6. Woods Hole Oceanographic Institution Report 67-66, A Compilation of Moored Current Meter Observations, by F. Webster and N. P. Fofonoff, November 1967.

7. Lilliefors, H. W., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," Journal of the American Statistical Association, v. 62, p. 399-402, 1967.

8. Pearson, E. S., The Selected Papers of E. S. Pearson, University of California Press, Berkeley and Los Angeles, 1966.

9. Brooks, C. E. F. and Carruthers, N., Handbook of Statistical Methods in Meteorology, M. O. 538, Her Majesty's Stationery Office, 1953.

10. Andrews, D. F. and others, Robust Estimates of Location; Survey and Advances, Princeton University Press, 1972.

11. Pearson, E. S. and Hartley, H. O., Biometrika Tables for Statisticians, v. 2, p. 75-89, University Printing House, Cambridge, 1972.

12. D'Agostino, R. and Pearson, E. S., "Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$," Biometrika, v. 60, p. 613-622, 1973.

13. Bowman, K. O., "Power of the Kurtosis Statistic, b_2, in Tests of Departures from Normality," Biometrika, v. 60, p. 623-628, 1973.

14. Wilk, M. V. and Gnanadesikan, R., "Probability Plotting Methods for the Analysis of Data," Biometrika, v. 55, p. 1-17, 1968.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department of Oceanography, Code 58 Naval Postgraduate School Monterey, California 93940	3
4.	Oceanographer of the Navy Hoffman Building No. 2 200 Stovall Alexandria, Virginia 22332	1
5.	Office of Naval Research Code 480 Arlington, Virginia 22217	1
6.	Dr. Robert E. Stevenson Scientific Liaison Office, ONR Scripps Institution of Oceanography La Jolla California 92037	1
7.	Library, Code 3330 Naval Oceanographic Office Washington, D. C. 20373	1
8.	SIO Library University of California, San Diego P.O. Box 2367 La Jolla, California 92037	1
9.	Department of Oceanography Library University of Washington Seattle, Washington 98105	1
10.	Department of Oceanography Library Oregon State University Corvallis, Oregon 97331	1
11.	Commanding Officer Fleet Numerical Weather Control Monterey, California 93940	1

12. Commanding Officer 1
Environmental Prediction Research Facility
Monterey, California 93940

13. Department of the Navy 1
Commander Oceanographic System Pacific
Box 1390
FPO, San Francisco 96610

14. Asst Professor Robert G. Paquette, Code 58Pa 5
Department of Oceanography
Naval Postgraduate School
Monterey, California 93940

15. Asst Professor Donald P. Gaver, Jr. Code 55Gv 1
Department of Operations Research and
Administrative Sciences
Naval Postgraduate School
Monterey, California 93940

16. Asst Professor Thomas D. Burnett, Code 55Za 1
Department of Operations Research and
Administrative Sciences
Naval Postgraduate School
Monterey, California 93940

Thesis 155501
M2768 Mahaffy
c.1 Investigation of the statistics of ocean current speeds.